

Auxiliary Likelihood-Based Approximate Bayesian Computation in State Space Models*

Gael M. Martin[†], Brendan P.M. McCabe[‡], David T. Frazier[§],
Worapree Maneesoonthorn[¶] and Christian P. Robert^{||}

January 9, 2017

Abstract

A new approach to inference in state space models is proposed, using approximate Bayesian computation (ABC). ABC avoids evaluation of an intractable likelihood by matching summary statistics computed from observed data with statistics computed from data simulated from the true process, based on parameter draws from the prior. Draws that produce a ‘match’ between observed and simulated summaries are retained, and used to estimate the inaccessible posterior; exact inference being feasible only if the statistics are sufficient. With no reduction to sufficiency being possible in the state space setting, we pursue summaries via the maximization of an auxiliary likelihood function. We derive conditions under which this auxiliary likelihood-based approach achieves Bayesian consistency and show that, in the limit, results yielded by the auxiliary maximum likelihood estimator are replicated by the auxiliary score. In multivariate parameter settings a separate treatment of each parameter dimension, based on integrated likelihood techniques, is advocated as a way of avoiding the curse of dimensionality associated with ABC methods. Three stochastic volatility models for which exact inference is either challenging or infeasible, are used for illustration.

Keywords: Likelihood-free methods, stochastic volatility models, Bayesian consistency, asymptotic sufficiency, unscented Kalman filter, α -stable distribution.

JEL Classification: C11, C22, C58

*This research has been supported by Australian Research Council Discovery Grant No. DP150101728. The authors would like to thank the Editor, an associate editor and two anonymous referees for very helpful and constructive comments on an earlier draft of the paper.

[†]Department of Econometrics and Business Statistics, Monash University, Australia. Corresponding author; email: gael.martin@monash.edu.

[‡]Management School, University of Liverpool, U.K.

[§]Department of Econometrics and Business Statistics, Monash University, Melbourne, Australia.

[¶]Melbourne Business School, University of Melbourne, Australia.

^{||}Université Paris-Dauphine, Centre de Recherche en Économie et Statistique, and University of Warwick.

1 Introduction

The application of Approximate Bayesian computation (ABC) (or likelihood-free inference) to models with intractable likelihoods has become increasingly prevalent of late, gaining attention in areas beyond the natural sciences in which it first featured. (See Beaumont, 2010, Csillery *et al.*, 2010; Marin *et al.*, 2011, Sisson and Fan, 2011 and Robert, 2015, for reviews.) The technique circumvents direct evaluation of the likelihood function by selecting parameter draws that yield pseudo data - as simulated from the assumed model - that matches the observed data, with the matching based on summary statistics. If such statistics are sufficient, and if an arbitrarily small tolerance is used in the matching, the selected draws can be used to produce a posterior distribution that is exact up to simulation error; otherwise, an estimate of the *partial* posterior - defined as the density of the unknown parameters conditional on the summary statistics - is the only possible outcome.

The choice of statistics for use within ABC, in addition to techniques for determining the matching criterion, are clearly of paramount importance, with much recent research having been devoted to devising ways of ensuring that the information content of the chosen set of statistics is maximized, in some sense; e.g. Joyce and Marjoram (2008), Wegmann *et al.* (2009), Blum (2010), Fearnhead and Prangle (2012) and Frazier *et al.* (2016). In this vein, Drovandi *et al.* (2011), Gleim and Pigorsch (2013), Creel and Kristensen (2015), Creel *et al.*, (2015) and Drovandi *et al.* (2015), produce statistics via an *auxiliary* model selected to approximate the features of the true data generating process. This approach mimics, in a Bayesian framework, the principle underlying the frequentist methods of indirect inference (Gouriéroux *et al.*, 1993, Smith, 1993) and efficient method of moments (Gallant and Tauchen, 1996) using, as it does, the approximating model to produce feasible inference about an intractable true model. Whilst the price paid for the approximation in the frequentist setting is a possible reduction in efficiency, the price paid in the Bayesian case is posterior inference that is conditioned on statistics that are not sufficient for the parameters of the true model, and which amounts to only partial inference as a consequence.

Our paper continues in this spirit, but with focus given to the application of auxiliary model-based ABC methods in the state space model (SSM) framework. Whilst ABC methods have been proposed in this setting (*inter alia*, Jasra *et al.*, 2010, Dean *et al.*, 2014, Martin *et al.*, 2014, Calvet and Czellar, 2015a, 2015b, Yildirim *et al.*, 2015), such methods use ABC principles (without summarization) to estimate either the likelihood function or the smoothed density of the states, with established techniques - for example, maximum likelihood or (particle) Markov chain Monte Carlo (PMCMC) - then being used to conduct inference on the static parameters themselves. (Jasra, 2015, provides an extensive review

of this literature, including existing theoretical results, as well as providing comprehensive computational insights.)

Our aim, in contrast, is to explore the use of ABC alone and as based on summarization via maximum likelihood estimation (MLE) of the parameters of an auxiliary model. Drawing on recent theoretical results on the properties of MLE in misspecified SSMs (Douc and Moulines, 2012) we provide a set of conditions that ensures that auxiliary likelihood-based ABC is *Bayesian consistent* in the state space setting, in the sense of producing draws that yield a degenerate distribution at the true vector of static parameters in the (sample size) limit. Use of maximum likelihood to estimate the auxiliary parameters also allows the concept of asymptotic sufficiency to be invoked, thereby ensuring that - for large samples at least - maximum information is extracted from the auxiliary likelihood in producing the summaries.

We also illustrate that to the order of accuracy that is relevant in establishing the theoretical properties of an ABC technique, a selection criterion based on the score of the auxiliary likelihood - evaluated at the maximum likelihood estimator (MLE) computed from the observed data - yields equivalent results to a criterion based directly on the MLE itself. This equivalence is shown to hold in both the exactly and over-identified cases, and independently of any positive definite weighting matrix used to define the two alternative distance measures, and implies that the proximity to asymptotic sufficiency yielded by using the auxiliary MLE in an ABC algorithm will be replicated by the use of the auxiliary score. Given the enormous gain in speed achieved by avoiding optimization of the auxiliary likelihood at each replication of ABC, this is a critical result from a computational perspective.

Finally, we briefly address the issue of dimensionality that impacts on ABC techniques in multiple parameter settings. (See Blum, 2010, Fearnhead and Prangle, 2012, Nott *et al.*, 2014 and Biau *et al.*, 2015). Specifically, we demonstrate numerically the improved accuracy that can be achieved by matching individual parameters via the corresponding scalar score of the integrated auxiliary likelihood, as an alternative to matching on the multi-dimensional score statistic as suggested, for example, in Drovandi *et al.* (2015).

We illustrate the proposed method in three classes of model for stochastic return volatility. Two of the classes exemplify the case where the transition densities in the state process have a representation that is either challenging to embed within an exact algorithm or is unavailable analytically. The third class of model illustrates the case where the conditional density of returns given the latent volatility is unavailable. Satisfaction of the sufficient conditions for Bayesian consistency of ABC (up to identification conditions) is demonstrated for one class. Examples from all three classes are then explored numerically, in artificial data scenarios, with consistent inference being confirmed. This being the first attempt made to formally verify the validity of auxiliary likelihood-based ABC techniques

in such complex settings, the results augur well for the future use of the method.

The paper proceeds as follows. In Section 2 we briefly summarize the basic principles of ABC as they would apply in a state space framework. In Section 3, we then proceed to demonstrate the theoretical properties of the auxiliary likelihood approach to ABC, including sufficient conditions for Bayesian consistency to hold, in this particular setting. The sense in which inference based on the auxiliary MLE is replicated by inference based on the auxiliary score is also described. In Section 4 we then consider the application of the auxiliary likelihood approach in the non-linear state space setting, using the three classes of latent volatility models for illustration. Numerical accuracy of the proposed method, as applied to data generated artificially from the Heston (1993) square root volatility model, is then assessed in Section 5.1. Existence of known (non-central chi-squared) transition densities means that the exact likelihood function/posterior distribution is available for the purpose of comparison. The accuracy of the auxiliary likelihood-based ABC posterior estimate is compared with: 1) an ABC estimate that uses a (weighted) Euclidean metric based on statistics that are sufficient for an observed autoregressive model of order one defined on the logarithmic squared returns; and 2) an ABC estimate that exploits the dimension-reduction technique of Fearnhead and Prangle (2012), applied to this latter set of summary statistics. The auxiliary likelihood-based method is shown to provide the most accurate estimate of the exact posterior in almost all cases documented. In Section 5.2 numerical evidence supports Bayesian consistency for the auxiliary-likelihood based method in all three SSMs investigated in the paper. In contrast, evidence for the consistency of various summary-statistic based ABC methods is mixed. Section 6 concludes. Technical proofs and certain computational details are included in appendices to the paper.

2 Auxiliary likelihood-based ABC in state space models

2.1 Outline of the basic approach

The aim of ABC is to produce draws from an approximation to the posterior distribution of a vector of unknowns, θ , given the T -dimensional vector of observed data $\mathbf{y} = (y_1, \dots, y_T)'$,

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta),$$

in the case where both the prior, $p(\theta)$, and the likelihood, $p(\mathbf{y}|\theta)$, can be simulated. These draws are used, in turn, to approximate posterior quantities of interest, including marginal posterior moments, marginal posterior distributions and predictive distributions. The simplest (accept/reject) form of the algorithm (Tavaré *et al.*, 1997, Pritchard, 1999) proceeds as follows:

Algorithm 1 ABC accept/reject algorithm

- 1: Simulate θ^i , $i = 1, 2, \dots, N$, from $p(\theta)$
- 2: Simulate $\mathbf{z}^i = (z_1^i, z_2^i, \dots, z_T^i)'$, $i = 1, 2, \dots, N$, from the likelihood, $p(\cdot|\theta^i)$
- 3: Select θ^i such that:

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}^i)\} \leq \varepsilon, \quad (1)$$

where $\eta(\cdot)$ is a (vector) statistic, $d\{\cdot\}$ is a distance criterion, and, given N , the tolerance level ε is chosen to be small.

The algorithm thus samples θ and \mathbf{z} from the joint posterior:

$$p_\varepsilon(\theta, \mathbf{z}|\eta(\mathbf{y})) = \frac{p(\theta)p(\mathbf{z}|\theta)\mathbb{I}_\varepsilon[\mathbf{z}]}{\int_{\Theta} \int_{\mathbf{Z}} p(\theta)p(\mathbf{z}|\theta)\mathbb{I}_\varepsilon[\mathbf{z}]d\mathbf{z}d\theta},$$

where $\mathbb{I}_\varepsilon[\mathbf{z}] := \mathbb{I}[d\{\eta(\mathbf{y}), \eta(\mathbf{z})\} \leq \varepsilon]$ is one if $d\{\eta(\mathbf{y}), \eta(\mathbf{z})\} \leq \varepsilon$ and zero else. Clearly, when $\eta(\cdot)$ is sufficient and ε arbitrarily small,

$$p_\varepsilon(\theta|\eta(\mathbf{y})) = \int_{\mathbf{Z}} p_\varepsilon(\theta, \mathbf{z}|\eta(\mathbf{y}))d\mathbf{z} \quad (2)$$

approximates the exact posterior, $p(\theta|\mathbf{y})$, and draws from $p_\varepsilon(\theta, \mathbf{z}|\eta(\mathbf{y}))$ can be used to estimate features of that exact posterior. In practice however, the complexity of the models to which ABC is applied, including in the state space setting, implies that sufficiency is unattainable. Hence, as $\varepsilon \rightarrow 0$ the draws can be used to estimate features of $p(\theta|\eta(\mathbf{y}))$ only.

Adaptations of the basic rejection scheme have involved post-sampling corrections of the draws using kernel methods (Beaumont *et al.*, 2002, Blum, 2010, Blum and François, 2010), or the insertion of Markov chain Monte Carlo (MCMC) and/or sequential Monte Carlo (SMC) steps (Marjoram *et al.*, 2003, Sisson *et al.*, 2007, Beaumont *et al.*, 2009, Toni *et al.*, 2009, and Wegmann *et al.*, 2009), to improve the accuracy with which $p(\theta|\eta(\mathbf{y}))$ is estimated, for any given number of draws. Focus is also given to choosing $\eta(\cdot)$ and/or $d\{\cdot\}$ so as to render $p(\theta|\eta(\mathbf{y}))$ a closer match to $p(\theta|\mathbf{y})$, in some sense; see Joyce and Marjoram (2008), Wegmann *et al.*, Blum (2010) and Fearnhead and Prangle (2012). In the latter vein, Drovandi *et al.* (2011) argue, in the context of a specific biological model, that the use of $\eta(\cdot)$ comprised of the MLEs of the parameters of a well-chosen approximating model, may yield posterior inference that is conditioned on a large portion of the information in the data and, hence, be close to exact inference based on $p(\theta|\mathbf{y})$. (See also Gleim and Pigorsch, 2013, Creel and Kristensen, 2015, Creel *et al.*, 2015, and Drovandi *et al.*, 2015, for related work.) It is the spirit of this approach that informs the current paper, but with our attention given to rendering the approach feasible in a general state space framework that encompasses a large number of the models that are of interest to practitioners.

2.2 ABC in state space models

The stochastic process $\{y_t\}_{t \geq 0}$ represents a stationary ergodic process taking values in a measure space (Y, \mathcal{F}_y) , with \mathcal{F}_y a Borel σ -field, specified according to an SSM that depends on an unobserved state process $\{x_t\}_{t \geq 0}$, taking values in a measure space (X, \mathcal{F}_x) , with \mathcal{F}_x a Borel σ -field. The SSM is parameterized by unknown parameters $\phi \in \Phi \subset \mathbb{R}^{d_\phi}$, and for each ϕ , the state and observed sequences are generated according to the following measurement and state equations:

$$y_t = b(x_t, w_t, \phi) \quad (3)$$

$$x_t = G_\phi(x_{t-1}) + \Sigma_\phi(x_{t-1})v_t, \quad (4)$$

where $\{w_t, v_t\}_{t \geq 0}$ are independent sequences of *i.i.d.* random variables, $b(\cdot), \Sigma_\phi(\cdot), G_\phi(\cdot)$ are known, potentially nonlinear functions depending on $\phi \in \Phi$, and the matrix $\Sigma_\phi(\cdot)$ is full-rank for all $\phi \in \Phi$, with Φ compact. For each $\phi \in \Phi$, we assume that equation (4) defines a transition density $p(x_t|x_{t-1}, \phi)$ and that equation (3) gives rise to the conditional density of the sequence $\{y_t\}_{t \geq 0}$. This allows us to state the measurement and transition densities respectively as:

$$p(y_t|x_t, \phi) \quad (5)$$

$$p(x_t|x_{t-1}, \phi). \quad (6)$$

Throughout the remainder, we denote the ‘true value’ generating $\{y_t\}_{t \geq 0}$ by $\phi_0 \in \Phi$, and denote by \mathbb{P} and \mathbb{E} the law and expectation of the stationary SSM associated with ϕ_0 .

The aim of the current paper is to use ABC principles to conduct inference about (5) and (6) through ϕ . Our particular focus is situations where at least one of (5) or (6) is analytically unavailable, or computationally challenging, such that exact MCMC- or SMC-based techniques are infeasible or, at the very least, computationally burdensome. Three such classes of examples are later explored in detail, with all examples related to the modelling of stochastic volatility for financial returns, and with one example highlighting the case of a continuous-time volatility process.

ABC methods can be implemented within these types of settings so long as simulation from (5) and (6) is straightforward and appropriate ‘summaries’ of the data are available. We conduct ABC-based inference by relying on the structure of the SSM in (3) and (4) to generate a simplified version of the SSM, which we then use to produce informative summary measures for use in ABC. Specifically, we consider a simplified and, hence, misspecified version of equations (3) and (4), where

$$y_t = a(x_t, \epsilon_t, \beta) \quad (7)$$

$$x_t = H_\beta(x_{t-1}) + S_\beta(x_{t-1})e_t, \quad (8)$$

with $\{\epsilon_t, e_t\}_{t \geq 0}$ independent sequences of *i.i.d.* random variables with well-behaved densities; $a(\cdot), S_\beta(\cdot), H_\beta(\cdot)$ known functions of unknown parameters β ; and $S_\beta(\cdot)$ full-rank for all $\beta \in \mathcal{B} \subset \mathbb{R}^{d_\beta}$. Together, we assume this specification ensures that $\{x_t\}_{t \geq 0}$ takes values in the measure space (X, \mathcal{F}_x) and leads to a known transition kernel $Q_\beta : X \times X \times \mathcal{B} \rightarrow [0, 1]$, which admits the known state-transition density $q_\beta(\cdot, \cdot) : X \times X \times \mathcal{B} \rightarrow \mathbb{R}_+$, and known conditional density $g_\beta : X \times Y \times \mathcal{B} \rightarrow \mathbb{R}_+$. That is, equations (7) and (8) imply that $x_t|x_{t-1} \sim q_\beta(x_t, x_{t-1})$ and $y_t|x_t \sim g_\beta(y_t, x_t)$, with both $q_\beta(\cdot, \cdot)$ and $g_\beta(\cdot, \cdot)$ analytically tractable.

Defining the parametric family of the above misspecified SSM as $\mathcal{G} := \{\beta \in \mathcal{B} : (q_\beta(x, x'), g_\beta(y, x))\}$, we maintain that there is no reason to assume $\mathbb{P} \in \mathcal{G}$. However, even if $\mathbb{P} \notin \mathcal{G}$, it will generally be the case that a well-chosen \mathcal{G} is capable of capturing many of the features associated with the DGP in equations (3) and (4). To this end, and in the spirit of indirect inference, we obtain summary statistics for ABC using the quasi-likelihood associated with the parametric family \mathcal{G} . Such a strategy requires defining the quasi-likelihood associated with the misspecified SSM, which, following Gouriéroux *et al.* (1993), amongst others, is hereafter referred to as the auxiliary likelihood. Defining $\chi(\cdot)$ to be an initial probability measure on (X, \mathcal{F}_x) , for $\mathbf{y}_m^T = (y_m, \dots, y_T)'$, we state the auxiliary likelihood for inference on β as

$$p_\chi(\mathbf{y}_m^T; \beta) = \int \cdots \int \chi(dx_m) g_\beta(y_m, x_m) \prod_{p=m+1}^T Q_\beta(x_{p-1}, dx_p) g_\beta(y_p, x_p).$$

From observations $\mathbf{y} = \mathbf{y}_1^T \equiv (y_1, \dots, y_T)'$, the auxiliary MLE can then be obtained as

$$\hat{\beta}(\mathbf{y}) = \arg \max_{\beta \in \mathcal{B}} L_a(\mathbf{y}; \beta); \quad L_a(\mathbf{y}; \beta) = \log(p_\chi(\mathbf{y}_1^T; \beta)). \quad (9)$$

Given $\eta(\mathbf{y}) = \hat{\beta}(\mathbf{y})$, ABC can then proceed via Algorithm 1.

We note that, in the above setting, the full set of unknowns constitutes the augmented vector $\theta = (\phi', \mathbf{x}'_c)'$ where, in the case when x_t evolves in continuous time, \mathbf{x}_c represents the infinite-dimensional vector comprising the continuum of unobserved states over the sample period. However, to fix ideas, we define $\theta = (\phi', \mathbf{x}')'$, where $\mathbf{x} = (x_1, x_2, \dots, x_T)'$ is the T -dimensional vector comprising the time t states for the T observation periods in the sample.¹ Implementation of the ABC algorithm thus involves simulating ϕ from the prior $p(\phi)$, followed by simulation of x_t via the process for the state, conditional on the draw of ϕ , and subsequent simulation of artificial data z_t conditional on the draws of ϕ and the state variable. Crucially, our attention is given to inference about ϕ only; hence, only draws of ϕ are retained (via the selection criterion) and those draws used to produce an estimate of the marginal posterior, $p(\phi|\mathbf{y})$. That is, from this point onwards, when we reference a

¹For example, in a continuous-time stochastic volatility model such values may be interpreted as end-of-day volatilities.

vector of summary statistics, $\eta(\mathbf{y})$, for instance, $\eta(\mathbf{y}) = \widehat{\beta}(\mathbf{y})$, it is the information content of that vector with respect to ϕ that is of importance, and the asymptotic behaviour of $p_\varepsilon(\phi|\eta(\mathbf{y}))$ with reference to the true ϕ_0 that is under question. Similarly, in the numerical illustration in Section 5.1, it is the proximity of the particular (kernel-based estimate of) $p_\varepsilon(\phi|\eta(\mathbf{y}))$ explored therein to the exact $p(\phi|\mathbf{y})$ that is documented. We comment briefly on state inference in Section 6.

3 Auxiliary likelihood-based ABC

3.1 ‘Approximate’ asymptotic sufficiency

ABC is predicated on the use of ‘informative’ summaries in its implementation, with a vector of sufficient statistics being the only form of summary that replicates the information content of the full sample, and with the Pitman-Koopman-Darmois Theorem establishing that sufficiency is attainable only for distributions that are members of the exponential family (EF). For the general SSM described by (5) and (6) for any t - and with our particular focus being cases where either density does not have an analytical representation - the joint distribution of \mathbf{y} will, almost by default, not be in the EF, and sufficiency reduction will therefore not be feasible.²

On the other hand, asymptotic Gaussianity of the MLE for the parameters of (5) and (6) implies (under regularity) that the MLE satisfies the factorization theorem and is thereby asymptotically sufficient for the parameters of that model. (See Cox and Hinkley, 1974, Chp. 9 for elucidation of this matter.) Denoting the log-likelihood function by $L(\mathbf{y}; \phi)$, maximizing $L(\mathbf{y}; \phi)$ with respect to ϕ yields $\widehat{\phi}$, which could, *in principle*, be used to define $\eta(\cdot)$ in an ABC algorithm. For large enough T (and for $\varepsilon \rightarrow 0$) the algorithm would thus produce draws from the exact posterior. Indeed, in arguments that mirror those adopted by Gouriéroux *et al.* (1993) and Gallant and Tauchen (1996) for the indirect inference and efficient method of moments estimators respectively, Gleim and Pigorsch (2013) demonstrate that if $\eta(\cdot)$ is chosen to be the MLE of an auxiliary model that nests (or ‘smoothly embeds’) the true model in some well-defined way, asymptotic sufficiency for the true parameters will still be achieved; see also Gouriéroux and Monfort (1995) on this point.

Of course, if the SSM in question is such that the exact likelihood is accessible, the

²Even the simplest SSMs, with all components available, generate moving average-like dependence in the data. The linear Gaussian SSM is the leading case, and for which simple computations lead to an analytical link between the signal-to-noise ratio and the lack of sufficiency associated with any finite set of statistics calculated from the observations. The crux of the problem is that information in the sample does not ‘accumulate’ in the way required for reduction to a sufficient set of statistics of dimension smaller than T to be feasible (see, for example, Anderson, 1958, Chp. 6). The essence of this problem would characterize any SSM nested in (5) and (6), simply due to the presence of measurement error.

model is likely to be tractable enough to preclude the need for treatment via ABC, with the primary goal of this paper being the presentation of ABC methods in SSMs for which exact methods are essentially infeasible. Further, the quest for asymptotic sufficiency via a *nesting* auxiliary model conflicts with the quest for an accurate non-parametric estimate of the posterior using the ABC draws, given that the dimension of the parameter set in the auxiliary model is, by construction, likely to be large. Hence, in practice, the appropriate goal in using the auxiliary likelihood approach to ABC in the SSM context is to define, via (7) and (8), a sensible *parsimonious* approximation to the true model in (5) and (6), for which the associated likelihood function can be evaluated with computational ease and speed. Heuristically, if the approximating model is ‘accurate enough’ as a representation of the true model, such an approach will yield, via the ABC algorithm, an estimate of the posterior distribution that is conditioned on a statistic that is ‘close to’ being asymptotically sufficient for ϕ . We certainly make no attempt in this paper to formalize this statement in any way. Nevertheless, we do view the notion of asymptotic sufficiency of the auxiliary MLE as being a intuitively compelling characteristic of the auxiliary likelihood-based approach to ABC, and the numerical results presented later provide some support for its importance in practice. More critically, however, pursuing the auxiliary likelihood route enables us to draw on regularity as it pertains to likelihood functions, and maximization thereof, to prove the Bayesian consistency of the resultant ABC posterior and, hence, the baseline accuracy of the inferences produced via this route.

3.2 Consistency of auxiliary likelihood-based ABC

For a given choice of auxiliary model in (7) and (8), with parameters $\beta \in \mathbf{B} \subset \mathbb{R}^{d_\beta}$, $d_\beta \geq d_\phi$, and sample log-likelihood function $L_a(\mathbf{y}; \beta)$ defined in (9), ABC can use as summary statistics for inference on ϕ the maximizers of $L_a(\cdot; \beta)$, based on \mathbf{y} and $\mathbf{z}(\phi^i)$, which we represent respectively by

$$\hat{\beta}(\mathbf{y}) = \arg \max_{\beta \in \mathbf{B}} L_a(\mathbf{y}; \beta) \text{ and } \hat{\beta}(\mathbf{z}(\phi^i)) = \arg \max_{\beta \in \mathbf{B}} L_a(\mathbf{z}(\phi^i); \beta).$$

Herein, $\mathbf{z}(\phi^i)$ is the *ith* vector of pseudo data, with the dependence of $\mathbf{z}(\phi^i)$ on the *ith* random draw ϕ^i from the prior $p(\phi)$ made explicit in the notation. Using $\hat{\beta}(\mathbf{y})$ and $\hat{\beta}(\mathbf{z}(\phi^i))$ as summary statistics, we can take as the distance criterion in (1),

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi^i))\} = \sqrt{\left[\hat{\beta}(\mathbf{y}) - \hat{\beta}(\mathbf{z}(\phi^i)) \right]' \mathbf{\Omega} \left[\hat{\beta}(\mathbf{y}) - \hat{\beta}(\mathbf{z}(\phi^i)) \right]}, \quad (10)$$

where $\mathbf{\Omega}$ is some positive definite matrix.

As noted above, with sufficiency and, hence, exact posterior inference via ABC, being an unachievable goal in the complex state space settings that we envisage here, we aim to establish conditions under which ABC attains a weaker - but no less important - form

of validity, namely Bayesian consistency. Under such conditions the investigator can be assured that, at the very least, with a large enough sample size the ABC posterior will concentrate on the true parameter vector and provide valid inference in that sense.

In the ABC setting, Bayesian consistency requires that as $T \rightarrow \infty$ and $\varepsilon \rightarrow 0$, the estimated posterior based on the selected draws from $p_\varepsilon(\phi|\eta(\mathbf{y}))$ concentrates around the true parameter value generating the data; see, for example, Frazier *et al.* (2016) and the references therein. With a slight abuse of terminology, from this point onwards we denote the ‘ABC posterior’ by $p_\varepsilon(\phi|\eta(\mathbf{y}))$, recognizing that the quantity produced via ABC is actually the kernel-based density estimate constructed from a given number of draws, N , from $p_\varepsilon(\phi|\eta(\mathbf{y}))$ as defined in (2).

To understand the intuition underlying Bayesian consistency of ABC based on $\eta(\mathbf{y}) = \widehat{\beta}(\mathbf{y})$, first define $Z \subseteq Y$ to be the space of simulated data $\mathbf{z}(\phi)$, generated according to the probability measure P_z^ϕ , and denote the prior measure of a set $A \subset \Phi$ by $\Pi(A)$. We also make it explicit from this point onwards that Bayesian consistency depends on simultaneous asymptotics regarding T and ε . To formalize this we consider ε as a T -dependent sequence, denoted by ε_T , where $\varepsilon_T \rightarrow 0$ as $T \rightarrow \infty$.

Heuristically, Bayesian consistency of ABC would then follow from the following sequence of arguments. First, under mild regularity conditions, as $T \rightarrow \infty$, the criterion in (10) should satisfy

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi^i))\} \xrightarrow{P} \sqrt{[\beta_0 - \mathbf{b}(\phi^i)]' \boldsymbol{\Omega} [\beta_0 - \mathbf{b}(\phi^i)]}, \quad (11)$$

where “ \xrightarrow{P} ” denotes convergence in probability, and where

$$\beta_0 = \arg \max_{\beta \in \mathbf{B}} \left\{ \text{plim}_{T \rightarrow \infty} L_a(\mathbf{y}; \beta)/T \right\}; \quad \mathbf{b}(\phi^i) = \arg \max_{\beta \in \mathbf{B}} \left\{ \text{plim}_{T \rightarrow \infty} L_a(\mathbf{z}(\phi^i); \beta)/T \right\}.$$

Secondly, under identification conditions, $\phi^i = \phi_0$ should be the only value that satisfies $\beta_0 = \mathbf{b}(\phi^i)$ and, as a consequence, the only value that satisfies

$$d\{\beta_0, \mathbf{b}(\phi^i)\} = \sqrt{[\beta_0 - \mathbf{b}(\phi^i)]' \boldsymbol{\Omega} [\beta_0 - \mathbf{b}(\phi^i)]} = 0. \quad (12)$$

Hence, as $T \rightarrow \infty$, for any $\varepsilon_T > 0$ such that $\Pi[\{\phi^i \in \Phi : d\{\beta_0, \mathbf{b}(\phi^i)\} \leq \varepsilon_T\}] > 0$, the only value of ϕ^i satisfying $d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi^i))\} \leq \varepsilon_T$ for all ε_T is $\phi^i = \phi_0$; therefore, if $\widehat{\beta}(\mathbf{y})$ is well-behaved, as $T \rightarrow \infty$, $\varepsilon_T \rightarrow 0$, the ABC algorithm will only select draws arbitrarily close to ϕ_0 . Put formally, the ABC posterior will be Bayesian consistent if, for any $\delta > 0$ and $A_\delta(\phi_0) := \{\phi \in \Phi : d\{\phi, \phi_0\} > \delta\}$,

$$\int_{A_\delta(\phi_0)} p_\varepsilon(\phi|\eta(\mathbf{y})) d\phi = \int_{A_\delta(\phi_0)} \frac{\int_Z 1 \left[d\{\widehat{\beta}(\mathbf{y}), \widehat{\beta}(\mathbf{z}(\phi))\} \leq \varepsilon_T \right] P_z^\phi(d\mathbf{z}) \Pi(d\phi)}{\int_\Phi \int_Z 1 \left[d\{\widehat{\beta}(\mathbf{y}), \widehat{\beta}(\mathbf{z}(\phi))\} \leq \varepsilon_T \right] P_z^\phi(d\mathbf{z}) \Pi(d\phi)} = o_P(1), \quad (13)$$

as $T \rightarrow \infty$ and $\varepsilon_T \rightarrow 0$.

Establishing (13) in this SSM setting requires $\eta(\mathbf{y}) = \widehat{\beta}(\mathbf{y}) \xrightarrow{P} \beta_0$ and, uniformly in ϕ^i , $\eta(\mathbf{z}(\phi^i)) = \widehat{\beta}(\mathbf{z}(\phi^i)) \xrightarrow{P} \mathbf{b}(\phi^i)$. It also requires continuity and injectivity of the so-called ‘binding function’ $\phi \mapsto \mathbf{b}(\cdot)$. Sufficient conditions to guarantee the convergence in (13) can be split into two sets: the first controls the convergence of sample quantities; the second set comprises identification conditions.

Assumption A:

- (A1) The parameter spaces $\mathcal{B} \subset \mathbb{R}^{d_\beta}$ and $\Phi \subset \mathbb{R}^{d_\phi}$ are compact.
- (A2) For any $\phi \in \Phi$, $z_t(\phi) \in Z \subseteq Y$, $\{z_t(\phi), x_t(\phi)\}_{t=1}^T$ is a stationary and ergodic process, with $(z_0(\phi), x_0(\phi))$ drawn in the stationary law.
- (A3) For $(x, x', \beta) \mapsto q_\beta(x, x')$ the density of the Markov transition kernel associated with the auxiliary model satisfies the following:
 - (A3.1) $(x, x', \beta) \mapsto q_\beta(x, x')$ is a positive continuous function on $X \times X \times \mathcal{B}$.
 - (A3.2) $\sup_{\beta \in \mathcal{B}} \sup_{(x, x') \in X \times X} q_\beta(x, x') < \infty$.
- (A4) The conditional density, $(y, x, \beta) \mapsto g_\beta(y, x)$, associated with the auxiliary model satisfies the following conditions:
 - (A4.1) For each $(x, y) \in X \times Y$, $(y, x, \beta) \mapsto g_\beta(y, x)$ is positive and continuous on $Y \times X \times \mathcal{B}$.
 - (A4.2) For any $\mathcal{K} \subset Y$, compact, and any $\beta \in \mathcal{B}$, $\lim_{|x| \rightarrow \infty} \sup_{y \in \mathcal{K}} \frac{g_\beta(y, x)}{\sup_{x' \in X} g_\beta(y, x')} = 0$.
 - (A4.3) For $z_0(\phi) \in Y$ as in (A2), $\mathbb{E}_\phi [\ln^+ \sup_{\beta \in \mathcal{B}} \sup_{x \in X} g_\beta(z_0(\phi), x)] < \infty$.
 - (A4.4) There exists a compact subset $\mathcal{D} \subset X$ such that, for $z_0(\phi) \in Y$ as in (A2), $\mathbb{E}_\phi [\ln^- \inf_{\beta \in \mathcal{B}} \inf_{x \in \mathcal{D}} g_\beta(z_0(\phi), x)] < \infty$.
- (A5) $L_\infty(\phi^i; \beta) := \text{plim}_{T \rightarrow \infty} (1/T) L_a(\mathbf{z}(\phi^i); \beta)$ has unique maximum $\mathbf{b}(\phi^i) = \arg \max_{\beta \in \mathcal{B}} L_\infty(\phi^i; \beta)$, where $\beta_0 = \mathbf{b}(\phi_0) = \arg \max_{\beta \in \mathcal{B}} L_\infty(\phi_0; \beta)$.

Assumption I:

- (I1) For $\Psi_\varepsilon := \{\phi \in \Phi : d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon\}$, some $D > 0$ and a constant $K > 0$, the prior satisfies $\Pi(\Psi_\varepsilon) \geq K\varepsilon^D$.
- (I2) The mapping $\phi \mapsto \mathbf{b}(\phi)$ is continuous and one-to-one.

(I3) For any $\phi \in \Phi$, there exist constants $\kappa, C, u_0 > 0$ such that, for some sequence $v_T \rightarrow \infty$ and all $0 < u < u_0 v_T$,

$$P_z^\phi \left[d \left\{ \widehat{\beta}(\mathbf{z}(\phi)), \mathbf{b}(\phi) \right\} > u \right] \leq C(\phi) u^{-\kappa} v_T^{-\kappa}, \text{ and } \int_{\Phi} C(\phi) \Pi(d\phi) < \infty.$$

Remark 1: Under correct specification of the model generating the data \mathbf{y} , Assumptions (A1)-(A5) ensure that $\sup_{\beta \in \mathbf{B}} |(1/T)L_a(\mathbf{y}; \beta) - L_\infty(\phi_0; \beta)| = o_P(1)$, for $L_\infty(\phi_0; \beta)$ defined in (A5), and that $\|\widehat{\beta}(\mathbf{y}) - \beta_0\| = o_P(1)$, for $\|\cdot\|$ the Euclidean norm. In addition, Assumptions (A1)-(A5) are enough to ensure that $\sup_{\phi^i \in \Phi} \|\widehat{\beta}(\mathbf{z}(\phi^i)) - \mathbf{b}(\phi^i)\| = o_P(1)$. The uniform convergence of $\widehat{\beta}(\mathbf{z}(\phi^i))$ to $\mathbf{b}(\phi^i)$ is crucial as it ensures that the simulated paths $\mathbf{z}(\phi^i)$, and the subsequent $\widehat{\beta}(\mathbf{z}(\phi^i))$, are well-behaved over Φ . Assumptions (I1)-(I3) ensure the required concentration of the ABC posterior on sets containing the truth, ϕ_0 . In particular, Assumption (I1) ensures that the prior used within ABC places sufficient mass on the truth, and (some version of) this assumption is standard in the analysis of Bayesian consistency. Assumption (I3) is a type of deviation control for the estimated auxiliary parameters, and allows us precise control over certain remainder terms in the posterior decomposition.

The following theorem formally establishes Bayesian consistency of the ABC posterior.

Theorem 1 *For all $\delta > 0$, if Assumptions (A) and (I) are satisfied, then, so long as $\varepsilon_T = o(1)$ is such that $\varepsilon_T^{D+\kappa} v_T^\kappa \rightarrow \infty$, and Ω is positive definite,*

$$\int_{A_\delta(\phi_0)} p_\varepsilon(\phi | \eta(\mathbf{y})) d\phi = o_P(1), \text{ for } \eta(\mathbf{y}) = \widehat{\beta}(\mathbf{y}), \text{ as } T \rightarrow \infty,$$

where $A_\delta(\phi_0) := \{\phi \in \Phi : d\{\phi, \phi_0\} > \delta\}$.

Remark 2: The distance in (10) essentially mimics the Wald criterion used in the indirect inference technique.³ Similar to the latter, in our Bayesian analyses, in which (10) is used to produce ABC draws, Ω can also be defined as the sandwich form of a variance-covariance estimator (Gleim and Pigorsch, 2013, and Drovandi *et al.*, 2015), or as the inverse of the (estimated) variance-covariance matrix for β , evaluated at $\widehat{\beta}(\mathbf{y})$ (Drovandi *et al.*, 2011). In these cases it is more useful to denote the weighting matrix by $\widehat{\Omega}(\mathbf{y}, \widehat{\beta}(\mathbf{y}))$ and Bayesian consistency then requires, in addition to Assumptions (A) and (I), $\|\widehat{\Omega}(\mathbf{y}, \widehat{\beta}(\mathbf{y})) - \Omega_\infty(\beta_0)\|_* \xrightarrow{P} 0$, for some positive definite $\Omega_\infty(\beta_0)$, where $\|\mathbf{W}\|_* = \sqrt{\text{Trace}(\mathbf{W}'\mathbf{W})}$ for \mathbf{W} an arbitrary $n \times m$ matrix.

Remark 3: The conditions underlying Theorem 1 are weaker than those considered in the ABC literature where either the asymptotic shape of the ABC posterior or the asymptotic

³In practice the implementation of indirect inference may involve the use of a simulated sample in the computation of $\widehat{\beta}(\mathbf{z}(\phi^i))$ that is a multiple of the size of the empirical sample.

behaviour of ABC point estimates, or both, is the focus; see, for example, Creel *et al.* (2015), Frazier *et al.* (2016) and Li and Fearnhead (2016). For instance, nothing about our conditions requires the summaries to satisfy a central limit theorem.

Remark 4: We have presented the conditions for consistency, and proven Theorem 1, for the specific setting which is the focus here, namely where both the true and auxiliary models are SSMs. The sufficient conditions to ensure $\eta(\mathbf{y}) = \widehat{\beta}(\mathbf{y}) \xrightarrow{P} \beta_0$, and, uniformly in ϕ^i , $\eta(\mathbf{z}(\phi^i)) = \widehat{\beta}(\mathbf{z}(\phi^i)) \xrightarrow{P} \mathbf{b}(\phi^i)$ - **(A1)** to **(A5)** - are based on the conditions invoked by Douc and Moulines (2012) to establish consistency of the MLE in misspecified SSMs. Whilst these authors use simple examples to illustrate their theory, in our ABC setting, in which the true data generating process is, by the very nature of the exercise, a challenging one, verification of these conditions will not always be feasible. Similarly, it would appear to be infeasible to verify **(I3)** analytically under the remaining maintained assumptions in the typical case in which $\widehat{\beta}(\mathbf{z}(\phi))$ is unavailable in closed form. Moreover, and in common to all simulation-based inference procedures, analytical verification of the injectivity condition in **(I2)** is infeasible as a general rule, and, hence, remains an open problem. Nevertheless, we do illustrate the verification of Assumptions **(A1)** to **(A5)** in one class of examples, and demonstrate numerically that Bayesian consistency is achieved in all three classes considered.

Remark 5: In the numerical experiments, the distance in (1) is replaced by

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi^i))\} = \sqrt{\left[\mathbf{S}(\mathbf{z}(\phi^i); \widehat{\beta}(\mathbf{y}))\right]' \boldsymbol{\Sigma} \left[\mathbf{S}(\mathbf{z}(\phi^i); \widehat{\beta}(\mathbf{y}))\right]}, \quad (14)$$

where

$$\mathbf{S}(\mathbf{z}(\phi^i); \beta) = T^{-1} \frac{\partial L_a(\mathbf{z}(\phi^i); \beta)}{\partial \beta} \quad (15)$$

is the (average) score of the auxiliary likelihood, where $\mathbf{S}(\mathbf{y}; \widehat{\beta}(\mathbf{y})) = 0$, and $\boldsymbol{\Sigma}$ denotes a positive definite weighting matrix which, if an estimated quantity, satisfies comparable conditions to those specified in Remark 2 for $\widehat{\boldsymbol{\Omega}}(\cdot)$. Implementation of ABC via (14) is faster (by orders of magnitude) than the approach based upon $\eta(\cdot) = \widehat{\beta}(\cdot)$, due to the fact that maximization of the auxiliary likelihood is required only once, in order to produce $\widehat{\beta}(\cdot)$ from the observed data \mathbf{y} . All other calculations involve simply the evaluation of $\mathbf{S}(\cdot; \widehat{\beta}(\mathbf{y}))$ at the simulated data, with a numerical differentiation technique invoked to specify $\mathbf{S}(\cdot; \widehat{\beta}(\mathbf{y}))$, when not known in closed form. Whilst we do not re-cast the formal conditions for consistency in terms of the auxiliary score, in Appendix B we do demonstrate informally that, under an additional identification condition, for $T \rightarrow \infty$ and $\varepsilon_T \rightarrow 0$, the score and MLE-based ABC selection criteria will yield equivalent draws of ϕ and, hence, equivalent estimates of $p(\theta|\mathbf{y})$.

4 Auxiliary likelihood-based ABC for three classes of latent volatility models

4.1 Overview

Given the critical role played by volatility in asset pricing, portfolio management and the calculation of risk measures, a large segment of the empirical finance literature has been devoted to the construction and analysis of volatility models. Three decades of empirical studies have demonstrated that the *constant* volatility feature of a geometric Brownian motion process for an asset price is inconsistent with both the observed time variation in return volatility and the non-Gaussian characteristics of empirical distributions of returns; see Bollerslev *et al.* (1992) for a review. Empirical regularities documented in the option pricing literature, most notably implied volatility ‘smiles’, are also viewed as evidence that asset prices deviate from the geometric Brownian motion assumption underlying the Black and Scholes (1973) option price; see Garcia *et al.* (2010) for a recent review.

In response to these now well-established empirical findings, many alternative time-varying volatility models have been proposed, with continuous-time stochastic volatility (SV) models - often augmented by random jump processes - being particularly prominent of late. This focus on the latter form of models is due, in part, to the availability of (semi-) closed-form option prices, with variants of the ‘square root’ SV model of Heston (1993) becoming the workhorse of the empirical option pricing literature. Given the challenging nature of the (non-central chi-squared) transitions in this model, Bayesian analyses of it have typically proceeded by invoking (Euler) discretizations for both the measurement and state processes and applying MCMC- or SMC-based techniques to that discretized model (e.g. Eraker, 2004, Forbes *et al.*, 2007, Broadie *et al.*, 2007, Johannes *et al.*, 2009). It has also featured in the indirect inference and efficient method of moments literatures, as a very consequence of the difficulty of evaluating the exact likelihood (e.g. Andersen, Benzoni and Lund, 2002, and Gallant and Tauchen, 2010). It is of interest, therefore, to assess the performance of the proposed ABC method in the context of this form of model, and this is the focus of Section 4.2.

In Sections 4.3 and 4.4 we then pursue two alternative volatility models in which the distinctly non-Gaussian features of the innovations to conditional returns are captured via the use of α -stable processes (see, e.g. Carr and Wu, 2003, and Lombradi and Calzolari, 2009). With the α -stable process not admitting a closed-form representation for the density function, models in which it appears present challenges for exact inference and are thus a prime candidate for analysis via ABC, in particular given that such processes *can* be simulated via the algorithm proposed in Chambers *et al.* (1976, 1987).

To facilitate the link between the general theoretical material presented thus far and the

specific examples to follow, we use the notation $\phi(\beta)$ to denote the vector of parameters characterizing the true (auxiliary) model in each case, despite the interpretation of the parameters obviously differing from case to case. We also use y_t to denote the observed measure in each example, x_t to denote the latent state and w_t and v_t to denote the measurement and state errors, as is consistent with the notation defined in (3) and (4).

4.2 Square root stochastic volatility

In this section we begin by assuming an observed (de-meaned) logarithmic return, r_t , with the square root model for the variance x_t ,

$$r_t = x_t^{1/2} \eta_t, \quad (16)$$

$$dx_t = (\phi_1 - \phi_2 x_t) dt + \phi_3 \sqrt{x_t} v_t, \quad (17)$$

where $v_t = dW_t$ is a Brownian increment, and η_t is defined as an *i.i.d.* random variable with zero mean and variance 1. We observe a discrete sequence of returns, and our goal is to conduct Bayesian inference on the parameters governing the dynamics of volatility. We restrict the structural parameters as $2\phi_1 \geq \phi_3^2$ to ensure positive volatility, and for some M, φ , we impose $M \geq \phi_3, \phi_1, \phi_2 \geq \varphi > 0$. With these restrictions, x_t is mean reverting and as $t \rightarrow \infty$, x_t approaches a steady state gamma distribution, with $E[x_t] = \phi_1/\phi_2$ and $\text{var}(x_t) = \phi_3^2 \phi_1 / 2\phi_2^2$. The conditional distribution function is non-central chi-square, $\chi^2(2cV_t; 2q+2, 2u)$, with $2q+2$ degrees of freedom and non-centrality parameter $2u$. The transition density for x_t , conditional on x_{t-1} , is thus

$$p(x_t | x_{t-1}, \phi) = c \exp(-u - v) \left(\frac{v}{u} \right)^{q/2} I_q(2(uv)^{1/2}), \quad (18)$$

where $c = 2\phi_2/\phi_3^2(1 - \exp(-\phi_2))$, $u = cx_{t-1} \exp(-\phi_2)$, $v = cx_t$, $q = \frac{2\phi_1}{\phi_3^2} - 1$, and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q .

With both the conditional density in (5) and the transition density in (6) being available for this model, likelihood-based inference is, in principle, feasible. For example, whilst we are not aware of any exact Bayesian inference having been conducted on the model in (16) and (17), the PMCMC techniques developed by Flury and Shephard (2011) and Pitt *et al.* (2012) for simpler volatility models may well be applicable. Further, in Section 5.1, in order to produce an exact comparator for the ABC posterior estimate for this model, we apply the non-linear filter of Ng *et al.* (2013) to evaluate the likelihood, and numerically normalize the exact posterior using deterministic numerical integration techniques. However, we do not propose the latter as a computationally attractive (or readily generalizable) competitor to the ABC approach, simply using it in a one-off exercise for the purpose of evaluation; and the performance of a PMCMC algorithm for an SSM with transitions as challenging as those given in (18) is as yet untested. Hence, we view the

application of ABC in this setting as an attractive option to explore, in particular given the ability to *simulate* the process via its exact representation as a composition of central chi-squared and Poisson distributions.

For convenience, we take squares and logarithms of the measurement equation, defining

$$y_t = \ln(r_t^2) = \ln(x_t) + w_t \quad (19)$$

$$dx_t = (\phi_1 - \phi_2 x_t)dt + \phi_3 \sqrt{x_t} v_t, \quad (20)$$

where

$$w_t = \ln(\eta_t^2) - \omega. \quad (21)$$

If we adopt the specific distributional assumption of Gaussianity for η_t , w_t is a mean-zero log-chi-squared random variable with variance $\sigma_w^2 = \pi^2/2$. The Gaussian assumption is not, however, essential for the verification of the sufficient conditions for consistency of ABC as applied to this model class. Rather, we require only that the first two moments of η_t are finite and that the density is bounded. We view (19) and (20) as the true data generating process under analysis and refer to it hereafter as the SV-SQ class of model. Note that the (exact) discretization of (20) would place (19) and (20) precisely in the form of (3) and (4) in Section 2.2.

To implement an auxiliary likelihood-based ABC algorithm, we adopt a Gaussian approximation for w_t in (19) and an Euler discretization for (20), yielding the approximating model,

$$y_t = \ln(x_t) + \epsilon_t \quad (22)$$

$$x_t = \beta_1 + \beta_2 x_{t-1} + \beta_3 \sqrt{x_{t-1}} e_t, \quad (23)$$

where $\epsilon_t \sim N(0, \sigma_w^2)$, e_t is a truncated Gaussian variable with lower bound, $e_t > \frac{-\beta_1}{\beta_3}$, and we define the auxiliary parameters as $\beta = (\beta_1, \beta_2, \beta_3)'$. Similar parameter restrictions to those imposed on the structural parameters ϕ are required of the elements of β : $M \geq \beta_1$, $\beta_3 \geq \varphi > 0$, $\varphi \leq \beta_2 \leq 1 - \varphi$, and $2\beta_1 \geq \beta_3^2$. The equations in (22) and (23) play the role of (7) and (8) respectively.

The non-linearities that characterize both (22) and (23) imply that an analytical evaluation of the auxiliary likelihood via the Kalman filter (KF) is not feasible. Therefore, we turn to the augmented unscented KF (AUKF) as an computationally efficient means of evaluating the $L_a(\mathbf{y}; \beta)$ and, hence, of producing the auxiliary MLE as the matching statistic within ABC. General pseudo code detailing implementation of the AUKF is given in Algorithm 2, with more detailed implementation instructions given in Appendix C.1. The precise form of the auxiliary likelihood function thus depends on both the first-order Euler discretization of the continuous-time state process *and* the particular specifications used to implement the AUKF. For the AUKF specification detailed in Appendix C.1, we state the following corollary, the proof of which is given in Appendix C.2:

Algorithm 2 General AUKF algorithm

- 1: Initialize the system in (22) and (23) with a matrix of sigma-points $X_{a,0}$ and a vector of fixed weights; see, Appendix A.3.1 for the definition of these sigma-points and weights;
 - 2: **while** $t \leq T$ **do**
 - 3: Propagate $X_{a,t-1}$ through (23) to obtain x_t sigma points for time t ;
 - 4: Using simple weighted sums of the x_t sigma points, generate the predicted mean and variance for x_t ;
 - 5: Use the predicted mean and variance to generate a new matrix of sigma points $X_{a,t}$;
 - 6: Propagate $X_{a,t}$ through (22) to obtain y_t sigma points for time t ;
 - 7: Using simple weighted sums of the y_t sigma points, generate the predicted mean and variance for y_t ;
 - 8: Use the predicted mean and variance to form a Gaussian conditional density for y_t ;
 - 9: Using the predicted mean and variance for y_t and KF up-dating, produce the filtered mean and variance for x_t , given the observation of y_t , and up-date the sigma points $X_{a,t}$ accordingly;
 - 10: Set $t = t + 1$;
 - 11: **end while**
 - 12: $L_a(\mathbf{y}; \beta)$ is the log-product of the increments in Step 8.
-

Corollary 1 *For the SV-SQ model in (19) and (20) and true value ϕ_0 , the model in (22) and (23), with auxiliary likelihood $L_a(\mathbf{y}; \beta)$ constructed via the AUKF filter, and with $\eta(\mathbf{y}) = \hat{\beta}(\mathbf{y})$, satisfies Assumption (A1)-(A5).*

Remark 6: Bayesian consistent inference for ϕ_0 also depends on the satisfaction of Assumption (I). Whilst (I1) is trivially satisfied via the specification of a sensible prior, Assumptions (I2)-(I3) in the SV-SQ model are not amenable to analytical investigation, or verification, given the nature of the auxiliary likelihood, as numerically evaluated using the AUKF, and the lack of a closed-form expression for the auxiliary MLE. This is, in fact, an illustration of the general point that there exists a tension between a choice of summaries for which one can analytically verify Assumptions (I2) and (I3), and a choice of more complicated summaries for which analytical verification is not feasible, but which yield more accurate ABC-based inference. To this end, we set our focus on the latter but remark that if one were willing to consider a simpler auxiliary model (and associated auxiliary MLE) then verification of Assumptions (I2) and (I3) may well be possible analytically. Note that we do (in effect) investigate the satisfaction of (I2)-(I3) for the SV-SQ model in Section 5.2, in which consistency of the auxiliary model-based method for all three examples is explored numerically. Such numerical exploration is indeed the *only* option available to us for the two models that follow, given that analytical verification of both sets of conditions, (A) and (I), is precluded due to the need (in part) to evaluate certain expectations under an α -stable law.

4.3 Conditionally α -stable returns with stochastic volatility

Let $\{X_t^{\alpha,\gamma}, t \in \mathbb{R}_+\}$ be an α -stable Lévy process with location $\mu = 0$, scale $\sigma = 1$, tail index $\alpha \in (1, 2)$, and skewness parameter $\gamma \in [-1, 1]$. Then X_t has independent and stationary increments $dX_t^{\alpha,\gamma}$ such that $dX_t^{\alpha,\gamma} \sim \mathcal{S}(\alpha, \gamma, 0, dt^{1/\alpha})$ and exhibits differing degrees of leptokurtosis and skewness depending on the values of α and γ . The process is also self-similar in that the distribution of an α -stable variable defined over any horizon has the same shape upon scaling. Critically however, the density function has no closed-form representation. (See Samorodnitsky and Taqqu, 1994, Chapter 7.)

Recently, several authors have used α -stable Lévy motion to model financial data. Notably, Carr and Wu (2003) model logarithmic returns on the S&P500 price index as α -stable, with a view to capturing the lack of ‘flattening’ of the implied volatility smile as option maturity increases. In brief, the infinite variance (for the log return) implied by this model violates the conditions for a Gaussian central limit theorem and, hence, fits with the phenomenon of a smile that persists. At the same time, however, with the lower bound imposed for γ , the conditional expectation of the index itself remains finite, thereby enabling meaningful European option prices to be defined. Whilst the detailed derivations in their paper pertain to the case in which volatility is constant, recognition of the need to incorporate stochastic volatility prompts the authors to propose (as a vehicle for future research) an extended model in which the Heston (1993) model in (17) is adopted for the variance, with closed-form option pricing still being feasible as a consequence.

Most importantly, with the focus in Carr and Wu (2003) being on the estimation of risk neutral parameters via calibration of the model with market option prices, the lack of analytical form for the density of X_t is not a hindrance for inference. However, any attempt to conduct likelihood-based inference (including exact Bayesian inference) on the objective counterpart of such a model using spot returns *would* encounter this hurdle, with the conditional density in (5) being unavailable; and that is where ABC provides a useful alternative.

With this empirical motivation in mind, we thus explore the application of ABC to the model

$$y_t = r_t = x_t^{1/\phi_4} w_t, \tag{24}$$

$$\ln x_t = \phi_1 + \phi_2 \ln x_{t-1} + \phi_3 v_t, \tag{25}$$

where $w_t \sim i.i.d. \mathcal{S}(\phi_4, -1, 0, dt = 1)$, v_t is an *i.i.d.* random variable (independent of w_t) with zero mean and variance 1, and to be consistent with our general notation, we denote α by ϕ_4 . Once again we assume discretely observed returns and, for the sake of illustration, work with a discrete-time autoregressive model for the logarithm of the variance, as given in (25). In particular this allows us to illustrate ABC using the following simple aux-

iliary model based on a first-order generalized autoregressive conditional heteroscedastic (GARCH(1,1)) model for the latent standard deviation,

$$y_t = r_t = x_t \epsilon_t, \quad (26)$$

$$x_t = \beta_1 + \beta_2 x_{t-1} |\epsilon_{t-1}| + \beta_3 x_{t-1}, \quad (27)$$

where $\epsilon_t \sim i.i.d. St(0, 1, \beta_4)$. That is, the measurement error in the auxiliary model is a standardized Student t random variable with degrees of freedom parameter β_4 . (See also Lombardi and Calzolari, 2009, and Garcia *et al.*, 2011, for the application of indirect inference to similar model scenarios.) The ARCH component of (27) is parameterized using absolute deviations (instead of squares) to mitigate numerical instabilities that can arise from extreme realizations of the α -stable distribution. Note that the model in (26) and (27) can be placed in the state space form given in (7) and (8) by defining $e_t = |\epsilon_{t-1}|$; but with the auxiliary likelihood function available in closed form in this case, the application of ABC is particularly straightforward and does not require filtering.⁴

4.4 Stochastic volatility with α -stable errors

An alternative approach to modelling the stylized features of financial returns is to consider a stochastic volatility model for returns in which an α -stable process drives the innovations to (log) volatility itself; see Lombardi and Calzolari (2009) once again. To that end, in this section we define the following model for the return,

$$\begin{aligned} r_t &= x_t^{1/2} w_t, \\ \ln x_t &= \phi_1 + \phi_2 \ln x_{t-1} + \phi_3 v_t, \end{aligned}$$

where $v_t \sim i.i.d. \mathcal{S}(\phi_4, -1, 0, dt = 1)$, and w_t is an *i.i.d.* random variable (independent of v_t) with zero mean and variance. With this particular specification it is the transition density in (6) that is unavailable, rendering exact likelihood-based inference infeasible. In the spirit of Lombardi and Calzolari we base ABC on a (conventional) GARCH(1,1) auxiliary model for the latent variance:

$$\begin{aligned} y_t &= r_t = x_t^{1/2} \epsilon_t, \\ x_t &= \beta_1 + \beta_2 x_{t-1} \epsilon_{t-1}^2 + \beta_3 x_{t-1}, \end{aligned}$$

in which case the computational burden of the ABC method is comparable to that in Section 4.3.

⁴Use of the square root volatility model in (25) would also of course be feasible, but the heteroscedastic nature of the variance model would demand an auxiliary model that reflected that feature, along the lines of (23), and hence, entail the use of filtering to evaluate the auxiliary likelihood.

5 Numerical assessment of the auxiliary likelihood-based ABC method

We undertake a series of numerical exercises in which the accuracy of the auxiliary likelihood-based approach to ABC is documented. The first set of exercises, in Section 5.1, uses the SV-SQ model in (19) and (20) as the example, with a Gaussian assumption adopted for the conditional distribution of returns. The auxiliary likelihood function of the approximating model (defined by (22) and (23)) is evaluated using the AUKF in the manner described above. Existence of known (non-central chi-squared) transition densities means that the exact likelihood function/posterior distribution is available for the purpose of comparison. We perform that evaluation using the non-linear grid-based filtering method of Ng *et al.* (2013). The accuracy of the auxiliary likelihood-based ABC posterior estimate, for a given finite sample size, is compared with: 1) an ABC estimate that uses a (weighted) Euclidean metric based on statistics that are sufficient for an observed autoregressive model of order one for the log squared returns; and 2) an ABC estimate that applies the approach of Fearnhead and Prangle (2012) to this set of summaries. We consider both the case where a single parameter (only) is unknown (and dimensionality thus plays no role), and the case where two, and then all three parameters of the model are unknown. A dimension reduction technique for the multi-parameter case, based on marginalization of the auxiliary likelihood is proposed, and shown to produce more accurate estimates overall of the exact marginals.

In Section 5.2 we then explore the large sample behaviour of the ABC posterior estimates for all three classes of stochastic volatility model. In particular, we illustrate that despite the fact that the full set of conditions for consistency for the auxiliary likelihood-based approach are not analytically verifiable (for any of the three examples), numerical evidence supports the presence of posterior concentration on the truth in all three cases. In contrast, the evidence in favour of consistency for summary statistic-based ABC estimates is mixed.⁵

5.1 Finite sample accuracy: the SQ-SV model

5.1.1 Data generation and computational details

For the purpose of this illustration we simulate artificially an ‘empirical’ sample of size T from the model in (19) and (20), with the parameters set to values that yield observations on both r_t and x_t that match the characteristics of (respectively) daily returns and daily values of realized volatility (constructed from 5 minute returns) for the S&P500 stock

⁵Results are produced using the GAUSS and MATLAB programming languages. Subroutines written in C are used to perform the integration of the auxiliary likelihood needed for the dimension reduction technique described in Section 5.1.2.

index over the 2003-2004 period: namely, $\phi_1 = 0.004$; $\phi_2 = 0.1$; $\phi_3 = 0.062$. This relatively calm period in the stock market is deliberately chosen as a reference point, as the inclusion of price and volatility jumps, and/or a non-Gaussian conditional distribution in the model would be an empirical necessity for any more volatile period, such as that witnessed during the 2008/2009 financial crisis, for example. The aim of this exercise being to assess the accuracy of the alternative ABC methods in a non-linear state space setting, it is important to have access to the exact posterior, and the SV-SQ model - without additional distributional complexities - enables this posterior to be accessed, via the deterministic non-linear filtering method of Ng *et al.* (2013). In brief, the method of Ng *et al.* represents the recursive filtering and prediction distributions used to define the exact likelihood function as the numerical solutions of integrals defined over the support of w_t in (19), with deterministic integration used to evaluate the relevant integrals, and the *exact* transitions in (20) used in the specification of the filtering and up-dating steps. Whilst lacking the general applicability of the ABC-based method proposed here, this deterministic filtering method is ideal for the particular model used in this illustration, and can be viewed as producing a very accurate estimate of the exact density, without any of the simulation error that would be associated with an MCMC-based comparator, for instance. We refer the reader to Ng *et al.* for more details of the technique; see also Kitagawa (1987).⁶ The likelihood function, evaluated via this method, is then multiplied by a uniform prior that imposes the restrictions: $0 < \phi_2 < 1$; $\phi_1, \phi_3 > 0$ and $2\phi_1 \geq \phi_3^2$, with ϕ_1 and ϕ_3 bounded above by 0.025 and 0.089 respectively. The three marginal posteriors are then produced via deterministic numerical integration (over the parameter space), with a very fine grid on ϕ being used to ensure accuracy. We report the posterior results for $1 - \phi_2$, where values of $1 - \phi_2$ close to unity signify a very persistent volatility process.

We compare the performance of the score-based technique with that of more conventional ABC methods based on summary statistics that may be deemed to be a sensible choice in this setting. For this purpose we propose a set of summary statistics that are sufficient (under Gaussianity) for an observable AR(1) process for the log of squared daily returns, $y_t = \ln(r_t^2)$, namely

$$s_1 = \sum_{t=2}^{T-1} y_t, \quad s_2 = \sum_{t=2}^{T-1} y_t^2, \quad s_3 = \sum_{t=2}^T y_t y_{t-1}, \quad s_4 = y_1 + y_T, \quad s_5 = y_1^2 + y_T^2. \quad (28)$$

⁶We note that the application of this filter in Ng *et al.* is to a non-parametric representation of the measurement error. In the current setting, in which w_t is specified parametrically, the known form of the distribution of w_t is used directly in the evaluation of the relevant integrals. We refer the reader to Section 2.2 of that paper for a full description of the algorithm. Preliminary experimentation with the number of grid points used in the deterministic integration was undertaken in order to ensure that the resulting estimate of the likelihood function/posterior stabilized, with 100 grid points underlying the final results documented here. As an additional check we also evaluated the exact (normalized) likelihood function using a bootstrap particle filter, based on 50,000 particle draws. The filtering-based estimate was indistinguishable from the grid-based estimate and, hence, is not reproduced here.

Two forms of distances are used. Firstly, we apply the conventional Euclidean distance, with each summary statistic also weighted by the inverse of the variance of the values of the statistic across the ABC draws. That is, we define

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi^i))\} = [\sum_{j=1}^5 (s_j^i - s_j^{obs})^2 / \text{var}(s_j)]^{1/2} \quad (29)$$

for ABC iteration $i = 1, 2, \dots, N$, where $\text{var}(s_j)$ is the variance (across i) of the s_j^i , and s_j^{obs} is the observed value of the j th statistic, $j = 1, 2, \dots, 5$. Secondly, we use a distance measure proposed in Fearnhead and Prangle (2012) which, as made explicit in Blum *et al.* (2013), is a form of dimension reduction method. We explain this briefly as follows. Given the vector of observations \mathbf{y} , the set of summary statistics in (28) are used to produce an estimate of $E(\phi_j|\mathbf{y})$, $j = 1, 2, 3$, which, in turn, is used as the summary statistic in a subsequent ABC algorithm. The steps of the Fearnhead and Prangle procedure (as modified for this context) for selection of the scalar parameter ϕ_j , $j = 1, 2, 3$, are as given in Algorithm 3.

Algorithm 3 FP ABC algorithm

- 1: Simulate ϕ^i , $i = 1, 2, \dots, N$, from $p(\phi)$
- 2: Simulate $\mathbf{x}^i = (x_1^i, x_2^i, \dots, x_T^i)'$ from (20) using the exact transitions, and pseudo data, \mathbf{z}^i using $p(\mathbf{z}|\mathbf{x}^i)$
- 3: Given \mathbf{z}^i , construct

$$\mathbf{s}^i = [s_1^i, s_2^i, s_3^i, s_4^i, s_5^i]' \quad (30)$$

- 4: For $\phi_j = (\phi_j^1, \phi_j^2, \dots, \phi_j^N)'$, $\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{s}^1 & \mathbf{s}^2 & \dots & \mathbf{s}^N \end{bmatrix}'$ and $\phi_j = E[\phi_j|\mathbf{Z}] + \mathbf{e} = \mathbf{X} [\alpha \quad \gamma']' + \mathbf{e}$, where $\mathbf{Z} = [\mathbf{z}^1, \mathbf{z}^2, \dots, \mathbf{z}^N]$ and γ is of dimension (5×1) , use OLS to estimate $E[\phi_j|\mathbf{Z}]$ as $\hat{E}[\phi_j|\mathbf{Z}] = \hat{\alpha} + [\mathbf{s}^1 \quad \mathbf{s}^2 \quad \dots \quad \mathbf{s}^N]' \hat{\gamma}$
- 5: For $\eta(\mathbf{z}^i) = \hat{E}(\phi_j|\mathbf{z}^i) = \hat{\alpha} + \mathbf{s}^{i'} \hat{\gamma}$ and $\eta(\mathbf{y}) = \hat{E}(\phi_j|\mathbf{y}) = \hat{\alpha} + \mathbf{s}^{obs'} \hat{\gamma}$, where \mathbf{s}^{obs} denotes the vector of summary statistics in (30) calculated from the vector of observed returns, use:

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}^i)\} = \left| \hat{E}(\phi_j|\mathbf{y}) - \hat{E}(\phi_j|\mathbf{z}^i) \right| = \left| \mathbf{s}^{i'} \hat{\gamma} - \mathbf{s}^{obs'} \hat{\gamma} \right| \quad (31)$$

as the selection criterion for ϕ_j .

The score-based method uses the distance measure in (14). The weighting matrix Σ is set equal to the Hessian-based estimate of the variance-covariance matrix of the (joint) MLE of β , evaluated at the MLE computed from the observed data, $\hat{\beta}(\mathbf{y})$. For the case where a single parameter only is unknown, the absolute value of the relevant scalar score is used to define (14). The 1% percentile of 50,000 ABC draws is used to determine the tolerance level.

5.1.2 Dimension reduction via an integrated likelihood technique

As highlighted by Blum (2010) (amongst others) the accuracy with which ABC draws estimate the so-called partial posterior, $p(\phi|\eta(\mathbf{y}))$, for any given tolerance ε and number of simulation draws N , will be less, the larger the dimension of $\eta(\mathbf{y})$. This ‘curse of dimensionality’ obtains even when the parameter ϕ is a scalar, and relates solely to the dimension of $\eta(\mathbf{y})$. As elaborated on further by Nott *et al.* (2014), this problem is exacerbated as the dimension of ϕ itself increases, firstly because an increase in the dimension of ϕ brings with it a concurrent need for an increase in the dimension of $\eta(\mathbf{y})$ and, secondly, because the need to estimate a multi-dimensional density (for ϕ) brings with it its own problems related to dimension.⁷ This type of inaccuracy is, of course, distinct from the inaccuracy that results from the use of summary statistics that are not sufficient for ϕ .

We explore here a dimension reduction technique that is particularly apt when there is a natural link between the elements of the true and auxiliary parameter vectors, and the dimensions of the two vectors are equivalent. These conditions are clearly satisfied for the model investigated in this particular numerical exercise, in which we produce the auxiliary model by discretization of the true latent diffusion, and $d_\beta = d_\phi$ as a consequence. In brief: let $\beta_{-j} = (\beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_{d_\phi})'$ be the $(d_\phi - 1)$ -dimensional parameter vector of auxiliary parameters (with $d_\phi = 3$ in this case), and $\mathbf{B}_{-j} \subset \mathbb{R}^{(d_\phi-1)}$ be the parameter space associated with β_{-j} . For $p(\beta_{-j}|\beta_j)$ the conditional prior probability of β_{-j} , define the integrated likelihood $L_a^I(\mathbf{y}; \beta_j)$ as

$$L_a^I(\mathbf{y}; \beta_j) = \int_{\mathbf{B}_{-j}} L_a(\mathbf{y}; \beta) p(\beta_{-j}|\beta_j) d\beta_{-j}. \quad (32)$$

For the given auxiliary model and conditional prior specification, $L_a^I(\mathbf{y}; \beta_j)$ can be used to obtain a convenient *scalar* summary statistic for use in estimating the marginal posterior $p(\phi_j|\mathbf{y})$ via ABC, using the integrated score,

$$S^I(\mathbf{z}(\phi); \hat{\beta}_j) = \frac{\partial \log (L_a^I(\mathbf{z}(\phi); \beta_j))}{\partial \beta_j} \Big|_{\beta_j = \hat{\beta}_j},$$

evaluated at $\hat{\beta}_j = \arg \max_{\beta_j} L_a^I(\mathbf{y}; \beta_j)$, where ϕ_j represents the true parameter that most closely matches the role played by β_j in the auxiliary model. If the marginal posteriors only are of interest, then all d_ϕ marginals can be estimated in this way, with d_ϕ applications of $(d_\phi - 1)$ -dimensional integration required at each step within ABC to produce the relevant score statistics. For the particular auxiliary model used here, the three integrated likelihoods are produced using a deterministic numerical method. If the *joint* posterior of ϕ were of interest, the sort of techniques advocated by Nott *et al.* (2014), amongst others, could be used to yield joint inference from the estimated marginal posteriors.

⁷See Blum *et al.* (2013) for further elaboration on the dimensionality issue in ABC and a review of current approaches for dealing with the problem.

5.1.3 Numerical results

We summarize accuracy by reporting the average, over 50 runs of ABC, of the root mean squared error (RMSE) of each ABC-based estimate of the exact (marginal) posterior for a given parameter, computed as:

$$RMSE = \sqrt{\frac{1}{G} \sum_{g=1}^G (\hat{p}_g - p_g)^2}, \quad (33)$$

where \hat{p}_g is the ordinate of the ABC density estimate and p_g the ordinate of the exact posterior density, at the g th grid-point used to produce the plots. The (average) RMSE associated with a given ABC method for any particular parameter is reported as a ratio to the RMSE of the (integrated) auxiliary score (‘Int Sc.’) method.

In order to abstract initially from the impact of dimensionality on the ABC methods, we report results in Panel A for each single parameter of the SV-SQ model, keeping the remaining two parameters fixed at their true values. In this case the auxiliary-likelihood method is based on the scalar score statistic, but the RMSE results are recorded in the row headed ‘Int score’. As is clear, for $1 - \phi_2$ the auxiliary score-based ABC method produces the most accurate estimate of the exact posterior of all comparators. In the case of ϕ_1 and ϕ_3 the summary statistic method (based on the Euclidean distance) yields the most accurate estimate, with the dimension reduction technique of Fearnhead and Prangle (2012) producing the least accurate posterior estimates for both parameters. The results recorded in Panels B to D highlight that when either two or three parameters are unknown the score-based ABC method produces the most accurate density estimates in all cases, with the integrated likelihood technique described in Section 5.1.2 yielding further accuracy improvements over the joint score (‘Jt Sc.’) methods in five out of the seven cases, auguring quite well for this particular approach to dimension reduction.

5.2 Large sample performance

5.2.1 Data generation and computational details

For all three examples outlined in Section 4 we now document numerically the extent to which the auxiliary likelihood-based ABC posteriors become increasingly concentrated (or otherwise) around the true parameters as the sample size increases. To this end, in Table 2 we report the average probability mass (over 50 runs of ABC) within a small interval around the true parameter, for $T = 500$ and 2000. Artificial ‘empirical’ data is generated from the SV-SQ model using the same parameter settings as detailed in Section 5.1.1. Generation from the other two models uses parameter settings that also yield empirically plausible data. Once again, as a means of comparison, summary statistic-based results are also produced, using both the Euclidean distance in (29) and the Fearnhead and Prangle

Table 1: Average RMSE of an estimated marginal posterior over 50 runs of ABC (each run using 50,000 replications, with 500 draws (1%) retained); recorded as a ratio to the (average) RMSE for the (integrated) ABC score method. ‘Sc.’ refers to the ABC method based on the score of the AUKF model; ‘SS’ refers to the ABC method based on a Euclidean distance for the summary statistics in (28); ‘FP’ refers to the Fearnhead and Prangle ABC method, based on the summary statistics in (28). For the single parameter case, the (single) score method is documented in the row denoted by ‘Int Sc.’, whilst in the multi-parameter case, there are results for both the joint (Jt) and integrated (Int) score methods. The bolded figure indicates the approximate posterior that is the most accurate in any particular instance. The sample size is $T = 500$.

ABC Meth.	Panel A One unknown			Panel B Two unknowns		Panel C Two unknowns		Panel D Three unknowns		
	ϕ_1	$1 - \phi_2$	ϕ_3	ϕ_1	$1 - \phi_2$	ϕ_2	ϕ_3	ϕ_1	$1 - \phi_2$	ϕ_3
Jt Sc.	-	-	-	1.689	1.613	0.873	1.843	1.652	0.408	1.015
Int Sc.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SS	0.529	4.926	0.673	2.310	3.685	3.625	2.229	2.441	1.850	1.036
FP	1.142	6.148	1.380	2.253	3.789	4.365	2.013	1.959	1.850	1.039

(2012) distance in (30). In order to highlight the dependence of posterior concentration on the particular choice of summaries, we define the statistics in (28) using both $y_t = \ln(r_t^2)$ and $y_t = r_t$, with results for the latter choice recorded in the rows denoted by SS (raw) and FP (raw). All relevant probabilities are estimated via rectangular integration of the ABC kernel density ordinates, with the boundaries of the interval used for a given parameter (recorded at the top of the table) determined by the grid used to numerically estimate the kernel density.

In order to reduce the computational burden, for the SV-SQ model we compute all probabilities for the (three) single unknown parameter cases only, and as based on 50,000 replications within each of the 50 ABC runs. For the other two models however, since the auxiliary models employed for both examples feature likelihood functions that are computationally simple, all parameters are estimated jointly. For both examples, we fix $\phi_1 = 0$, leaving three free parameters, ϕ_2 to ϕ_4 . Further, as guided by the theoretical results in Frazier *et al.* (2016), for these two models the quantile used to select draws is allowed to decline as T increases. With 250 draws retained for the purpose of density estimation this means that 55,902 and 447,214 replications (for each of the 50 draws) are used to produce the $T = 500$ and $T = 2000$ results respectively.

5.2.2 Numerical results

The results in Panel A (for $T = 500$) show that for the SQ-SV model the score-based method produces superior results - in terms of the extent of the probability mass around the truth - for ϕ_2 and ϕ_3 , with there being little to choose between the score-based estimate and the (equally accurate) two FP-based estimates in the case of ϕ_1 . Importantly, when the sample size increases the score-based estimate displays clear evidence of increased concentration around the true parameter value, and with the score based method clearly dominating all alternative (summary statistic-based) methods for $T = 2000$. Indeed, the tendency towards increased concentration is not uniform across all summary statistic methods, with the methods that exploit summaries constructed from the raw returns not exhibiting concentration for all three parameters.

The results in Panels B and C similarly illustrate the overall superiority of the score-based method and its consistency property, providing numerical evidence that the identification conditions hold in these particular examples. For the four alternative (summary statistic-based) methods however, the numerical evidence of posterior concentration is not uniform. It is also interesting to note that for the tail index parameter ϕ_4 in the SV model with conditionally stable returns (Panel B), the ABC method based on the raw data summary statistics ($y_t = r_t$) performs much better than the method based on the transformed data $y_t = \ln(r_t^2)$. In short, the logarithmic transformation of the conditionally stable returns yields summaries that are unable to estimate (via ABC) the true index parameter with any accuracy, no matter what the sample size, and no matter what the nature of the distance measure used. In contrast, when the α -stable distribution characterizes the errors in the volatility equation (Panel C) only the FP method applied to the summary statistics constructed from the log squared returns exhibits this extreme lack of concentration about the true value of ϕ_4 (for both sample sizes). Nevertheless, none of the summary statistic methods show a uniform tendency to concentrate further as the sample size increases for this particular example.

6 Conclusions and discussion

This paper has explored the application of approximate Bayesian computation in the state space setting, in which auxiliary likelihood functions are used to generate the matching statistics. Bayesian consistency of the auxiliary likelihood-based method has been established, under regularity conditions that exploit the state space structure of the auxiliary model. Theoretical verification of (certain of) these conditions has been established for one model class, with numerical evidence of posterior concentration produced for three model types. The idea of tackling the dimensionality issue via an integrated likelihood

Table 2: Posterior mass in given intervals around the true parameters, averaged over 50 runs of ABC. ‘Score’ refers to the ABC method based on the score of the relevant auxiliary model; ‘SS’ refers to the ABC method based on a Euclidean distance for the summary statistics in (28); ‘FP’ refers to the Fearnhead and Prangle ABC method, based on the summary statistics in (28); SS (raw) and FP (raw) refer, respectively to the SS and FP results but as based on $y_t = r_t$ rather than $y_t = \ln(r_t^2)$. The bolded figure indicates the largest (average) posterior mass for each case. Results in Panel A are for the SV-SQ model described in Section 4.2, with one parameter at a time treated as unknown. The results in Panels B and C are for the models in Section 4.3 and 4.4 respectively, with all three parameters for each model treated as unknown and the (joint) auxiliary score method used to produce the results recorded in the row headed ‘Score’.

$T = 500$				$T = 2000$		
Panel A: SV-SQ Model						
	ϕ_1	$1 - \phi_2$	ϕ_3	ϕ_1	$1 - \phi_2$	ϕ_3
True:	0.004	0.9	0.062	0.004	0.9	0.062
Interval:	(0.003,0.005)	(0.88,0.92)	(0.052,0.072)	(0.003,0.005)	(0.88,0.92)	(0.052,0.072)
Score	0.90	0.88	0.44	1.00	0.94	0.85
SS	0.78	0.44	0.26	0.99	0.83	0.57
FP	0.92	0.26	0.10	0.99	0.82	0.71
SS (raw)	0.84	0.28	0.44	0.78	0.24	0.87
FP (raw)	0.89	0.76	0.41	0.91	0.61	0.87
Panel B: Stable returns with SV ($\phi_1 = 0$)						
	ϕ_2	ϕ_3	ϕ_4	ϕ_2	ϕ_3	ϕ_4
True:	0.9	0.36	1.8	0.9	0.36	1.8
Interval:	(0.75,0.99)	(0.25,0.45)	(1.65,1.95)	(0.75,0.99)	(0.25,0.45)	(1.65,1.95)
Score	0.91	0.66	0.87	1.00	0.82	0.99
SS	0.15	0.57	0.15	0.44	0.61	0.00
FP	0.70	0.57	0.00	0.75	0.62	0.00
SS (raw)	0.78	0.48	0.91	0.78	0.47	1.00
FP (raw)	0.78	0.47	0.92	0.78	0.47	0.99
Panel C: SV with stable errors ($\phi_1 = 0$)						
	ϕ_2	ϕ_3	ϕ_4	ϕ_2	ϕ_3	ϕ_4
True:	0.9	0.06	1.8	0.9	0.06	1.8
Interval:	(0.75,0.99)	(0.03,0.09)	(1.65,1.95)	(0.75,0.99)	(0.03,0.09)	(1.65,1.95)
Score	0.84	0.72	0.39	0.96	0.74	0.58
SS	0.78	0.67	0.35	0.78	0.66	0.35
FP	0.76	0.70	0.09	0.78	0.73	0.02
SS (raw)	0.78	0.67	0.35	0.78	0.66	0.35
FP (raw)	0.78	0.67	0.36	0.79	0.67	0.33

approach has also been proposed. The approach has been shown to yield some benefits in the particular numerical example explored in the paper. However, a much more comprehensive analysis of different non-linear settings (and auxiliary models) would be required for a definitive conclusion to be drawn about the trade-off between the gain to be had from marginalization and the loss that may stem from integrating over an inaccurate auxiliary likelihood.

Indeed, the most important challenge that remains, as is common to the related frequentist techniques of indirect inference and efficient methods of moments, is the specification of a computationally efficient and accurate approximation. Given the additional need for parsimony, in order to minimize the number of statistics used in the matching exercise, the principle of aiming for a large nesting model, with a view to attaining full asymptotic sufficiency, is not an attractive one. We have illustrated the use of parsimonious approximating models. The relative success of this approach in the particular examples considered, certainly in comparison with methods based on other more ad hoc choices of summary statistics, augurs well for the success of auxiliary likelihood-based methods in the state space setting.

Finally, we note that despite the focus of this paper being on inference about the static parameters in the state space model, there is nothing to preclude marginal inference on the states being conducted, at a second stage. Specifically, conditional on the (accepted) draws used to estimate $p(\phi|\mathbf{y})$, existing filtering and smoothing methods (including the recent methods, referenced earlier, that exploit ABC at the filtering/smoothing level) could be used to yield draws of the states, and (marginal) smoothed posteriors for the states produced via the usual averaging arguments. With the asymptotic properties of both approaches established (under relevant conditions), of particular interest would be a comparison of both the finite sample accuracy and the computational burden of the hybrid ABC-based methods that have appeared in the literature, with that of the method proposed herein, in which $p(\phi|\mathbf{y})$ is targeted more directly via ABC principles alone.

References

- [1] Andersen, T.G., Benzoni, L. and Lund, J. 2002. An Empirical Investigation of Continuous-Time Equity Return Models, *The Journal of Finance*, 57, 1239–1284.
- [2] Anderson, T.W. 1958. *The Statistical Analysis of Time Series*, John Wiley and Sons.
- [3] Beaumont, M.A. 2010. Approximate Bayesian Computation in Evolution and Ecology, *Annual Review of Ecology, Evolution, and Systematic*, 41, 379-406.
- [4] Beaumont, M.A., Cornuet, J-M., Marin, J-M. and Robert, C.P. 2009. Adaptive Approximate Bayesian Computation, *Biometrika* 96, 983-990.

- [5] Beaumont, M.A., Zhang, W. and Balding, D.J. 2002. Approximate Bayesian Computation in Population Genetics, *Genetics* 162, 2025-2035.
- [6] Biau, G., Cérou, F. and Guyader, A. 2015. New insights into Approximate Bayesian Computation. *Ann. Inst. H. Poincaré Probab. Statist.*, 51, 376-403.
- [7] Black, F. and Scholes, M. 1973. The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81, 637-659.
- [8] Blum, M.G.B. 2010. Approximate Bayesian Computation: a Nonparametric Perspective, *Journal of the American Statistical Association* 105, 1178-1187.
- [9] Blum, M.G.B. and François, O. 2010. Non-linear Regression Models for Approximate Bayesian Computation, *Statistics and Computing* 20, 63-73.
- [10] Blum, M.G.B., Nunes, M.A., Prangle, D. and Sisson, S.A. 2013. A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation, *Statistical Science*, 28, 189-208.
- [11] Bollerslev, T., Chou, R.Y. and Kroner, K.F. 1992. ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence, *Journal of Econometrics*, 52, 5-59.
- [12] Broadie, M., Chernov, M. and Johannes, M. 2007. Model Specification and Risk Premia: Evidence from Futures Options, *The Journal of Finance*, LXII: 1453-1490.
- [13] Calvet, C. and Czellar, V. 2015a. Accurate Methods for Approximate Bayesian Computation Filtering. *Journal of Financial Econometrics* 13, 798-838.
- [14] Calvet, C. and Czellar, V. 2015b. Through the Looking Glass: Indirect Inference via Simple Equilibria. *Journal of Econometrics* 185, 343-358.
- [15] Carr, P. and Wu, L. 2003. The Finite Moment Log Stable Process and Option Pricing. *Journal of Finance* LVIII, 753-777.
- [16] Chambers, J.M., Mallows, C. and Stuck, B.W., 1976. A method for simulating stable random variables, *Journal of the American Statistical Association*, 71, 340-344.
- [17] Chambers, J.M., Mallows, C. and Stuck, B.W., 1987. Correction to: A method for simulating stable random variables. *Journal of the American Statistical Association*, 82, 704.
- [18] Cornuet, J-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J-M., Balding, D.J., Guillemand, T. and Estoup, A. 2008. Inferring Population History with DIY ABC: a User-friendly Approach to Approximate Bayesian Computation, *Bioinformatics* 24, 2713-2719.

- [19] Cox, D.R. and Hinkley, D.V. 1974. *Theoretical Statistics*, Chapman and Hall, London.
- [20] Creel, M. and Kristensen, D. 2015. ABC of SV: Limited Information Likelihood Inference in Stochastic Volatility Jump-Diffusion Models, *Journal of Empirical Finance* 31, 85-108.
- [21] Creel, M., Gao, J., Hong, H. and Kristensen, D. 2015. Bayesian Indirect Inference and the ABC of GMM. <http://arxiv.org/abs/1512.07385>.
- [22] Dean, T., Singh, S., Jasra, A. and Peters, G. 2014. Parameter Inference for Hidden Markov Models with Intractable Likelihoods, *Scand. J. Statist.* 41, 970-987.
- [23] Douc, R. and Moulines, E. 2012. Asymptotic Properties of the Maximum Likelihood Estimation in Misspecified Hidden Markov Models, *Ann. Statist.* 40, 2697-2732.
- [24] Drovandi, C.C., Pettitt, A.N. and Faddy, M.J. 2011. Approximate Bayesian Computation Using Indirect Inference, *JRSS(C)*, 60 1 - 21.
- [25] Drovandi, C.C., Pettitt, A.N. and Lee, A. 2015. Bayesian Indirect Inference Using a Parametric Auxiliary Model. *Statistical Science*, Vol. 30, No. 1, 72-95.
- [26] Eraker, B. 2004. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *The Journal of Finance*, LIX: 1367-1403.
- [27] Fearnhead, P. and Prangle, D. 2012. Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic Approximate Bayesian Computation. *Journal of the Royal Statistical Society, Series B.* 74: 419-474.
- [28] Flury, T. and Shephard, N. 2011. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models, *Econometric Theory*, 27(5), pp. 933-956.
- [29] Forbes C.S., Martin, G.M. and Wright J. 2007. Inference for a Class of Stochastic Volatility Models Using Option and Spot Prices: Application of a Bivariate Kalman Filter, *Econometric Reviews, Special Issue on Bayesian Dynamic Econometrics*, 26: 387-418.
- [30] Frazier D.T., Martin G.M., Robert, C.P. and Rousseau, J. 2016. Asymptotic Properties of Approximate Bayesian Computation, <https://arxiv.org/abs/1607.06903v2>.
- [31] Garcia, R., Ghysels, E. and Renault, E. 2010. The Econometrics of Option Pricing, *Handbook of Financial Econometrics*, Vol 1. (eds Ait-Sahalia, Y. and Hansen, L.), North Holland.

- [32] Garcia R., Renault, E. and Veredas, D. 2011. Estimation of stable distributions with indirect inference. *Journal of Econometrics*, 161(3): 325-337.
- [33] Gallant, A.R. and Tauchen, G. 1996. Which Moments to Match, *Econometric Theory* 12, 657-681.
- [34] Gallant, A.R. and Tauchen, G. 2010. Simulated Score Methods and Indirect Inference for Continuous-time Models, *Handbook of Financial Econometrics*, Volume 1, Chapter 8, Elsevier.
- [35] Gleim, A. and Pigorsch, C. 2013. Approximate Bayesian Computation with Indirect Summary Statistics. Draft paper: <http://ect-pigorsch.mee.uni-bonn.de/data/research/papers/>.
- [36] Gouriéroux, C. and Monfort, A. 1995. *Statistics and Econometric Models*. CUP.
- [37] Gouriéroux, C., Monfort, A. and Renault, E. 1993. Indirect Inference, *Journal of Applied Econometrics*, 85, S85-S118.
- [38] Heston, S.L. 1993. A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies*, 6, 327-343.
- [39] Jasra, A. 2015. Approximate Bayesian Computation for a Class of Time Series Models. *International Statistical Review* 83, 405-435.
- [40] Jasra, A, Singh, S, Martin, J. and McCoy, E. 2010. Filtering via Approximate Bayesian Computation. *Statistics and Computing* 22, 1223-1237.
- [41] Johannes, M., Polson, N.G. and Stroud, J.R. 2009. Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices, *Review of Financial Studies*, 22: 2759-2799.
- [42] Joyce, P. and Marjoram, P. 2008. Approximately Sufficient Statistics and Bayesian Computation. *Statistical applications in genetics and molecular biology*, 7, 1-16.
- [43] Julier, S.J., Uhlmann, J.K. and Durrant-Whyte, H.F. 1995. A New Approach for Filtering Nonlinear Systems. *Proceedings of the American Control Conference*, 1628-1632.
- [44] Julier, S.J., Uhlmann, J.K. and Durrant-Whyte, H.F. 2000. A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators, *IEEE Transactions on Automatic Control* 45, 477-481.

- [45] Kitagawa, G. 1987. Non-Gaussian State Space Modeling of Nonstationary Time Series. *Journal of the American Statistical Association* 76, 1032-1064.
- [46] Li, W. and Fearnhead, P. 2016. On the Asymptotic Efficiency of ABC Estimators, <https://arxiv.org/abs/1506.03481v3>.
- [47] Lombardi, M.J. and Calzolari, G. 2009. Indirect Estimation of α -stable Stochastic Volatility Models. *Computational Statistics and Data Analysis* 53 2298-2308.
- [48] Marin, J-M., Pudlo, P., Robert C. and Ryder, R. 2011. Approximate Bayesian Computation Methods. *Statistics and Computing* 21, 289–291.
- [49] Marjoram, P., Molitor, J., Plagnol, V. and Tavaré, S. 2003. Markov Chain Monte Carlo Without Likelihoods, *Proceedings of the National Academie of Science USA* 100, 15324-15328.
- [50] Martin, J. S., Jasra, A., Singh, S. S., Whiteley, N., Del Moral, P. and McCoy, E. 2014. Approximate Bayesian Computation for Smoothing, *Stoch. Anal. Appl.* 32, 397-422.
- [51] Ng, J., Forbes, C.S., Martin, G.M. and McCabe, B.P.M. 2013. Non-parametric Estimation of Forecast Distributions in Non-Gaussian, Non-linear State Space Models, *International Journal of Forecasting* 29, 411-430
- [52] Nott D., Fan, Y., Marshall, L. and Sisson, S. 2014. Approximate Bayesian Computation and Bayes Linear Analysis: Towards High-dimensional ABC, *Journal of Computational and Graphical Statistics*, 23, 65-86.
- [53] Pitt, M.K., dos Santos Silva, R., Giordani, P. and Kohn, R. 2012. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter, *Journal of Econometrics*, 171(2), pp. 134-151.
- [54] Pritchard, J.K., Seilstad, M.T., Perez-Lezaun, A. and Feldman, M.W. 1999. Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites, *Molecular Biology and Evolution* 16 1791-1798.
- [55] Robert, C.P. 2015. Approximate Bayesian Computation: a Survey on Recent Results, *Proceedings, MCQMC2014*.
- [56] Samoradnitsky, G. and Taqqu, M.S. 1994. *Stable non-Gaussian random processes: stochastic models with infinite variance*. Vol. 1. CRC press.
- [57] Sisson S. and Fan, Y. 2011. Likelihood-free Markov Chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (Eds. Brooks, Gelman, Jones, Meng). Chapman and Hall/CRC Press.

- [58] Sisson, S., Fan, Y. and Tanaka, M. 2007. Sequential Monte Carlo without Likelihoods, *Proceedings of the National Academy of Science USA* 104, 1760-1765.
- [59] Smith, A.A. 1993. Estimating Non-linear Time Series Models Using Vector Autoregressions: Two Approaches, *Journal of Applied Econometrics* 8, 63-84.
- [60] Tavaré, S., Balding, D.J., Griffiths, R.C. and Donnelly, P. 1997. Inferring Coalescence Times from DNA Sequence Data, *Genetics* 145, 505-518.
- [61] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. 2009. Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems, *JRSS (Interface)* 6, 187-202.
- [62] Wegmann, D., Leuenberger, C. and Excoffier, L. 2009. Efficient Approximate Bayesian Computation Coupled with Markov chain Monte Carlo with Likelihood, *Genetics* 182, 1207-1218.
- [63] Yildirim, S, Singh, S.S, Dean, T. and Jasra, A. 2015. Parameter Estimation in Hidden Markov Models with Intractable Likelihoods via Sequential Monte Carlo. *JCGS*, 24, 846-865.

A Proof of Theorem 1

Before beginning, let us first set notation used in the remainder of this appendix. Let $(\Omega_y, \mathcal{F}_y, \mathbb{P})$ be the fundamental space for the observed data, such that, for all $\omega \in \Omega_y$, $y_t = y_t(\omega) \in Y$, with Y the state space of the observed data. For any $\phi \in \Phi$ denote by $(\Omega_z, \mathcal{F}_z, P_z^\phi)$ the fundamental space of the simulated data, with which, for all $\omega_z \in \Omega_z$, $z_t(\phi) = z_t(\omega_z, \phi) \in Z \subseteq Y$. Furthermore, we assume the model is correctly specified so that for some $\phi_0 \in \Phi$, $P_z^{\phi_0} = \mathbb{P}$. Denote by $\Pi(A)$ the prior measure of $A \subset \Phi$. We index the tolerance ε by the sample size T to denote its eventual dependence on this value, so that $\varepsilon_T = \varepsilon(T) = o(1)$. C denotes a positive arbitrary constant. For two sequences, a_T, b_T , we say that $a_T \gg b_T$ if a_T is larger, ‘in order’, than b_T , and $a_T \asymp b_T$ if a_T is ‘of the same order’ as b_T . Lastly, we remind the reader that the ABC posterior measure is given by

$$\Pi_\varepsilon(A|\hat{\beta}(\mathbf{y})) := \Pi(A|d\{\hat{\beta}(\mathbf{y}), \hat{\beta}(\mathbf{z})\} \leq \varepsilon_T) = \int_A p_\varepsilon(\phi|\hat{\beta}(\mathbf{y}))d\phi.$$

To simplify notation, in what follows we use $\hat{\beta} := \hat{\beta}(\mathbf{y})$ and $\hat{\beta}(\phi) := \hat{\beta}(\mathbf{z}(\phi))$. The proof is broken into three parts. First, we demonstrate that the distance

$$d\{\eta(\mathbf{y}), \eta(\mathbf{z}(\phi))\} = d\{\hat{\beta}, \hat{\beta}(\phi)\} \equiv \sqrt{(\hat{\beta} - \hat{\beta}(\phi))' \boldsymbol{\Omega} (\hat{\beta} - \hat{\beta}(\phi))}$$

converges uniformly in ϕ to $d\{\beta_0, \mathbf{b}(\phi)\}$. Second, we show that the only value for which $d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T$ as $\varepsilon_T \rightarrow 0$ and $T \rightarrow \infty$ is ϕ_0 . Lastly, using pieces one and two we show that all posterior mass concentrates asymptotically on $\{\phi \in \Phi : d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T\}$.

A.1 Part 1

First, we demonstrate the uniform convergence of $d\{\hat{\beta}, \hat{\beta}(\phi)\}$. The triangle inequality yields

$$d\{\hat{\beta}, \hat{\beta}(\phi)\} \leq d\{\hat{\beta}, \mathbf{b}(\phi)\} + d\{\mathbf{b}(\phi), \hat{\beta}(\phi)\}.$$

Consider $d\{\hat{\beta}, \mathbf{b}(\phi)\}$ and note that,

$$d\{\hat{\beta}, \mathbf{b}(\phi)\} \leq d\{\hat{\beta}, \beta_0\} + d\{\beta_0, \mathbf{b}(\phi)\}.$$

From Theorem 2 and Proposition 3 of Douc and Moulines (2012), it can be shown that $\hat{\beta} \xrightarrow{P} \beta_0$ under Assumptions **(A1)**-**(A5)**. Consistency of $\hat{\beta}$ for β_0 implies

$$d\{\hat{\beta}, \mathbf{b}(\phi)\} \leq o_P(1) + d\{\beta_0, \mathbf{b}(\phi)\},$$

and so

$$d\{\hat{\beta}, \hat{\beta}(\phi)\} \leq d\{\hat{\beta}, \mathbf{b}(\phi)\} + d\{\mathbf{b}(\phi), \hat{\beta}(\phi)\} \leq d\{\beta_0, \mathbf{b}(\phi)\} + d\{\mathbf{b}(\phi), \hat{\beta}(\phi)\} + o_P(1).$$

By definition,

$$d\{\mathbf{b}(\phi), \hat{\beta}(\phi)\} \leq \sup_{\phi \in \Phi} d\{\mathbf{b}(\phi), \hat{\beta}(\phi)\}.$$

The RHS of the above is $o_P(1)$ if $\sup_{\phi \in \Phi} \|\hat{\beta}(\phi) - \mathbf{b}(\phi)\| \xrightarrow{P} 0$. Using Proposition 10, part (iii) in Douc and Moulines, and under Assumptions **(A1)**-**(A5)**, uniform convergence of $L_a(\mathbf{z}(\phi); \beta)/T$ over both β and ϕ can be shown, so that, for any $\delta = o(1)$, as $T \rightarrow \infty$

$$\sup_{\phi \in \Phi} \sup_{\|\beta - \beta_0\| \leq \delta} |L_a(\mathbf{z}(\phi); \beta)/T - L_\infty(\phi; \beta)| \xrightarrow{P} 0, \quad (34)$$

and by **(A5)** the limit criterion $L_\infty(\phi; \beta)$ has unique maximizer, with respect to β , $\mathbf{b}(\phi)$. Using these two facts, we now show, under the maintained assumptions,

$$\sup_{\phi \in \Phi} \|\hat{\beta}(\phi) - \mathbf{b}(\phi)\| = o_P(1). \quad (35)$$

Define the following terms:

$$\begin{aligned} \tilde{Q}(\phi; \beta) &= L_a(\mathbf{z}(\phi); \beta)/T - L_\infty(\phi; \mathbf{b}(\phi)), \\ \tilde{Q}_\infty(\phi; \beta) &= L_\infty(\phi; \beta) - L_\infty(\phi; \mathbf{b}(\phi)). \end{aligned}$$

By **(A5)** and compactness of \mathcal{B} , for all $\delta > 0$, if $\|\widehat{\beta}(\phi) - \mathbf{b}(\phi)\| > \delta$, there exists $\epsilon(\delta) > 0$, such that

$$\left| \widetilde{Q}_\infty(\phi; \widehat{\beta}(\phi)) \right| > \epsilon(\delta).$$

From here note that

$$\Pr \left(\sup_{\phi \in \Phi} \|\widehat{\beta}(\phi) - \mathbf{b}(\phi)\| > \delta \right) \leq \Pr \left(\sup_{\phi \in \Phi} \left| \widetilde{Q}_\infty(\phi; \widehat{\beta}(\phi)) \right| > \epsilon(\delta) \right).$$

Equation (35) follows if $\sup_{\phi \in \Phi} \|\widetilde{Q}_\infty(\phi; \widehat{\beta}(\phi))\| = o_P(1)$. Uniformly in ϕ ,

$$\begin{aligned} \left| \widetilde{Q}_\infty(\phi; \widehat{\beta}(\phi)) \right| &\leq \left| \widetilde{Q}_\infty(\phi; \widehat{\beta}(\phi)) - \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| + \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| \\ &= \left| L_\infty(\phi; \widehat{\beta}(\phi)) - L_a(\mathbf{z}(\phi); \widehat{\beta}(\phi))/T \right| + \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| \\ &\leq \sup_{\beta \in \mathbf{B}} |L_\infty(\phi; \beta) - L_a(\mathbf{z}(\phi); \beta)/T| + \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| \\ &\leq o_P(1) + \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right|. \end{aligned} \tag{36}$$

The first inequality follows from the triangle inequality, the second from the definition of $\widetilde{Q}_\infty(\phi; \beta)$, $\widetilde{Q}(\phi; \beta)$, the third from the definition of sup, and the last from (34). From (36), the result follows if

$$\sup_{\phi \in \Theta} \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| = o_P(1).$$

By the definition of $\widehat{\beta}(\phi)$, uniformly in ϕ ,

$$\begin{aligned} \left| \widetilde{Q}(\phi; \widehat{\beta}(\phi)) \right| &\leq \inf_{\beta \in \mathbf{B}} \left| \widetilde{Q}(\phi; \beta) \right| + o_P(1) \\ &\leq \inf_{\beta \in \mathbf{B}} \left| \widetilde{Q}(\phi; \beta) - \widetilde{Q}_\infty(\phi; \beta) \right| + \inf_{\beta \in \mathbf{B}} \left| \widetilde{Q}_\infty(\phi; \beta) \right| + o_P(1) \\ &\leq \sup_{\beta \in \mathbf{B}} |L_a(\mathbf{z}(\phi); \beta)/T - L_\infty(\phi; \beta)| + 0 + o_P(1) \\ &\leq o_P(1). \end{aligned} \tag{37}$$

Combining equations (36) and (37) yields $\sup_{\phi \in \Phi} \|\widehat{\beta}(\phi) - \mathbf{b}(\phi)\| = o_P(1)$ and so, uniformly in ϕ ,

$$d \left\{ \widehat{\beta}, \widehat{\beta}(\phi) \right\} \xrightarrow{P} d \left\{ \beta_0, \mathbf{b}(\phi) \right\}.$$

A.2 Part 2

The second portion of the proof demonstrates that, as $T \rightarrow \infty$ and $\varepsilon_T \rightarrow 0$, the only value of ϕ which the ABC algorithm selects is $\phi = \phi_0$. From the definition of the algorithm and the triangle inequality, for T large enough,

$$\begin{aligned} d \left\{ \widehat{\beta}, \widehat{\beta}(\phi) \right\} &\leq d \left\{ \widehat{\beta}, \beta_0 \right\} + d \left\{ \mathbf{b}(\phi), \widehat{\beta}(\phi) \right\} + d \left\{ \beta_0, \mathbf{b}(\phi) \right\} \\ &\leq \varepsilon_T/3 + \varepsilon_T/3 + d \left\{ \beta_0, \mathbf{b}(\phi) \right\} \end{aligned}$$

where the second inequality follows from the uniform convergence in equation (35) and $\widehat{\beta} = \beta_0 + o_P(1)$. A draw from the prior $p(\phi)$, will then be accepted, when T is large enough, if

$$d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T/3.$$

By Assumption **(I2)**, the only value of ϕ such that $d\{\mathbf{b}(\phi), \beta_0\} \leq \varepsilon_T/3$ as $\varepsilon_T \rightarrow 0$ is $\phi = \phi_0$.

A.3 Part 3

Given the previous two pieces, we can now show that as $T \rightarrow \infty$ and $\varepsilon_T \rightarrow 0$, for any $\delta > 0$ and $A_\delta(\phi_0) := \{\phi \in \Phi : d\{\phi, \phi_0\} > \delta\}$,

$$\Pi\left(A_\delta(\phi_0) | d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T\right) = \frac{\int_{A_\delta(\phi_0)} P_z^\phi(d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T) \Pi(d\phi)}{\int_{\Phi} P_z^\phi(d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T) \Pi(d\phi)} = o_P(1).$$

For $\Omega_\varepsilon = \{\mathbf{y} : d\{\widehat{\beta}, \beta_0\} \leq \varepsilon_T/3\}$, by **(A1)**-**(A5)** we have that $\mathbb{P}(\Omega_\varepsilon) = 1 + o(1)$, and $\mathbf{y} \in \Omega_\varepsilon$ with probability one. Now, consider the set

$$A_\varepsilon(\delta') := \left\{(\mathbf{z}, \phi) : \{d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T\} \cap \{d\{\beta_0, \mathbf{b}(\phi)\} > \delta'\}\right\}$$

For all $(\mathbf{z}, \phi) \in A_\varepsilon(\delta')$, we have, by the triangle inequality:

$$\delta' < d\{\beta_0, \mathbf{b}(\phi)\} \leq d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} + d\{\widehat{\beta}(\phi), \widehat{\beta}\} + d\{\widehat{\beta}, \beta_0\}.$$

For T large enough, by the results in Part 1 of the proof, $d\{\widehat{\beta}, \beta_0\} \leq \varepsilon_T/3$ and so, using the above inequality, we have

$$\delta' - (4/3)\varepsilon_T < d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\}.$$

Note that, for $\delta' \geq (5/3)\varepsilon_T$,

$$\begin{aligned} \Pr[A_\varepsilon(\delta')] &\leq \int_{\Phi} P_z^\phi(d\{\beta_0, \mathbf{b}(\phi)\} > \delta') \Pi(d\phi) \\ &\leq \int_{\Phi} P_z^\phi\left(d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} > \delta' - (4/3)\varepsilon_T\right) \Pi(d\phi) = o_P(1), \end{aligned} \quad (38)$$

by the results in Part 2 of the proof. From equation (38), we can conclude, for $\delta' = 4/3\varepsilon_T + s$ and $s \geq \varepsilon_T/3$,

$$\Pi\left(d\{\beta_0, \mathbf{b}(\phi)\} > \delta' | d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T\right) \leq \frac{\int_{\Phi} P_z^\phi\left(d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} > s\right) \Pi(d\phi)}{\int_{\Phi} P_z^\phi\left(d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T\right) \Pi(d\phi)}. \quad (39)$$

Using the above bound, the result follows if the RHS of equation (39) is $o_P(1)$.

First, focus on the denominator in (39). By the triangle inequality

$$\begin{aligned} d\{\widehat{\beta}, \widehat{\beta}(\phi)\} &\leq d\{\widehat{\beta}, \beta_0\} + d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} + d\{\beta_0, \mathbf{b}(\phi)\} \\ &\leq \varepsilon_T/3 + \varepsilon_T/3 + d\{\beta_0, \mathbf{b}(\phi)\} \end{aligned}$$

by Part 2 of the proof. Then, for any $\phi \in \Phi$ with $d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T/3$

$$\begin{aligned} \int_{\Phi} P_z^{\phi} \left(d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T \right) \Pi(d\phi) &\geq \int_{d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T/3} P_z^{\phi} \left(d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} \leq \varepsilon_T/3 \right) \Pi(d\phi) \\ &\geq \frac{\Pi[d\{\beta_0, \mathbf{b}(\phi)\} \leq \varepsilon_T/3]}{2} + o(1) \\ &\geq \frac{K\varepsilon_T^D}{2} + o(1), \end{aligned} \tag{40}$$

where the second inequality follows from the uniform convergence in equation (35), obtained in Part 1 of the proof, and the dominated convergence theorem; the last inequality follows by Assumption **(I1)**. Therefore, using equation (40) within equation (39) yields

$$\Pi \left(d\{\beta_0, \mathbf{b}(\phi)\} > \delta' | d\{\widehat{\beta}, \widehat{\beta}(\phi)\} \leq \varepsilon_T \right) \leq C \left[\int_{\Phi} P_z^{\phi} \left(d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} > s \right) \Pi(d\phi) \right] \varepsilon_T^{-D}. \tag{41}$$

The uniform convergence in (35) does not fully control the prior deviations. To ensure that the RHS of (41) is $o_P(1)$ we use Assumption **(I3)**, to conclude that

$$\left[\int_{\Phi} P_z^{\phi} \left(d\{\mathbf{b}(\phi), \widehat{\beta}(\phi)\} > s \right) \Pi(d\phi) \right] \varepsilon_T^{-D} \leq C \frac{\varepsilon_T^{-D}}{s^{\kappa} v_T^{\kappa}}.$$

Taking $s \asymp \varepsilon_T$, the RHS of (41) is $o_P(1)$ if $\varepsilon_T^{\kappa+D} v_T^{\kappa} \rightarrow \infty$, which is satisfied so long as $\varepsilon_T \gg v_T^{\frac{\kappa}{\kappa+D}}$. The result now follows from the arbitrary choice of δ' , and the continuity and injectivity of the map $\phi \mapsto \mathbf{b}(\phi)$. ■

B Auxiliary score-based ABC

Given the computational benefits of replacing the auxiliary MLE with the score, as the matching statistic, it is of interest to ascertain whether selection based on (14) will yield identical draws of ϕ to selection based on (10), at least in the appropriate limiting sense. If so, then the property of Bayesian consistency formally proven for the case of the MLE would hold, by default, for the score-based posterior estimate under appropriate regularity and identification conditions.

For any auxiliary likelihood (satisfying identification and regularity conditions) with unknown parameter vector β , we expand the (scaled) score function in (15), evaluated at $\widehat{\beta}(\mathbf{y})$, around the point $\widehat{\beta}(\mathbf{z}(\phi^i))$,

$$\mathbf{S}(\mathbf{z}(\phi^i); \widehat{\beta}(\mathbf{y})) = \mathbf{S}(\mathbf{z}(\phi^i); \widehat{\beta}(\mathbf{z}(\phi^i))) + \mathbf{D} \left[\widehat{\beta}(\mathbf{y}) - \widehat{\beta}(\mathbf{z}(\phi^i)) \right] = \mathbf{D} \left[\widehat{\beta}(\mathbf{y}) - \widehat{\beta}(\mathbf{z}(\phi^i)) \right], \tag{42}$$

where

$$\mathbf{D} = T^{-1} \frac{\partial^2 L_a(\mathbf{z}(\phi^i); \tilde{\beta}(\mathbf{z}(\phi^i)))}{\partial \beta \partial \beta'} \quad (43)$$

and $\tilde{\beta}(\mathbf{z}(\phi^i))$ denotes an (unknown, and coordinate-specific) intermediate value between $\hat{\beta}(\mathbf{y})$ and $\hat{\beta}(\mathbf{z}(\phi^i))$. Hence, using (42), the criterion in (14) becomes

$$\begin{aligned} & \sqrt{\left[\mathbf{S}(\mathbf{z}(\phi^i); \hat{\beta}(\mathbf{y})) \right]' \boldsymbol{\Sigma} \left[\mathbf{S}(\mathbf{z}(\phi^i); \hat{\beta}(\mathbf{y})) \right]} \\ &= \sqrt{\left[\hat{\beta}(\mathbf{y}) - \hat{\beta}(\mathbf{z}(\phi^i)) \right]' \mathbf{D}' \boldsymbol{\Sigma} \mathbf{D} \left[\hat{\beta}(\mathbf{y}) - \hat{\beta}(\mathbf{z}(\phi^i)) \right]} \leq \varepsilon_T. \end{aligned} \quad (44)$$

Subject to standard conditions regarding the second derivatives of the auxiliary likelihood, the matrix \mathbf{D} in (43) will be of full rank for $\hat{\beta}(\mathbf{z}(\phi^i))$ close to $\hat{\beta}(\mathbf{y})$. As a consequence, and given the positive definiteness of $\boldsymbol{\Sigma}$, $\mathbf{D}' \boldsymbol{\Sigma} \mathbf{D}$ will be a positive definite matrix that is some function of ϕ^i . Hence, whilst for any $\varepsilon_T > 0$, the presence of \mathbf{D} affects selection of ϕ^i , as $\varepsilon_T \rightarrow 0$, ϕ^i will be selected via (44) if and only if $\hat{\beta}(\mathbf{y})$ and $\hat{\beta}(\mathbf{z}(\phi^i))$ are equal. That is, the draws produced via the score-based criterion will be equivalent to those produced by using the auxiliary MLE itself as the matching statistic. Clearly, a *precise* statement on this equivalence requires a more rigorous discussion of the conditions required for identification, given that this equivalence need not hold for at least two reasons: one, for $\hat{\beta}(\mathbf{y})$ the auxiliary MLE, and $\hat{\beta}(\mathbf{z}(\phi))$, the zero of $\mathbf{S}(\mathbf{z}(\phi^i); \beta)$, there is no guarantee that $\hat{\beta}(\mathbf{z}(\phi_0))$ will converge to $\beta_0 = \text{plim}_{T \rightarrow \infty} \hat{\beta}(\mathbf{y})$ without further restrictions; two, there is no reason to believe that, even at $\phi^i = \phi_0$, the zero of $\mathbf{S}(\mathbf{z}(\phi^i); \beta)$ is unique. For brevity, and given the somewhat heuristic nature of this discussion, we do not consider such matters further.

Of course, in practice ABC is implemented with $\varepsilon_T > 0$, at which point the two ABC criteria will produce different draws. However, for the types of models entertained in this paper, preliminary investigation has assured us that the difference between the ABC estimates of the posteriors yielded by the alternative criteria is negligible for small enough ε_T . Hence, in the numerical section we operate solely with the score-based approach as the computationally feasible method of extracting both consistency and approximate asymptotic sufficiency in the state space setting.

C SV-SQ Example: AUKF Algorithm and Proof of Corollary 1

C.1 Detailed Implementation of the AUKF

Given the assumed invariance (over time) of both e_t and v_t in (22) and (23) respectively, the sigma points needed to implement the AUKF are determined as:

$$e^1 = E(e_t); \quad e^2 = E(e_t) + a_e \sqrt{\text{var}(e_t)}; \quad e^3 = E(e_t) - b_e \sqrt{\text{var}(e_t)}$$

and

$$v^1 = E(v_t); \quad v^2 = E(v_t) + a_v \sqrt{\text{var}(v_t)}; \quad v^3 = E(v_t) - b_v \sqrt{\text{var}(v_t)}$$

respectively, and propagated at each t through the relevant non-linear transformations, $h_t(\cdot)$ and $k_t(\cdot)$. The values a_e , b_e , a_v and b_v are chosen according to the assumed distribution of e_t and v_t , with a Gaussian assumption for both variables yielding values of $a_e = b_e = a_v = b_v = \sqrt{3}$ as being ‘optimal’. Different choices of these values are used to reflect higher-order distributional information and thereby improve the accuracy with which the mean and variance of the non-linear transformations are estimated; see Julier *et al.* (1995; 2000) for more details. Restricted supports are also managed via appropriate truncation of the sigma points. The same principles are applied to produce the mean and variance of the time varying state x_t , except that the sigma points need to be recalculated at each time t to reflect the up-dated mean and variance of x_t as each new value of y_t is realized.

In summary, the steps of the AUKF applied to evaluate the likelihood function of (22) and (23) are as follows:

1. Use the (assumed) marginal mean and variance of x_t , along with the invariant mean and variance of v_t and e_t respectively, to create the (3×7) matrix of augmented sigma points for $t = 0$, $X_{a,0}$, as follows. Define:

$$E(X_{a,0}) = \begin{bmatrix} E(x_t) \\ E(v_t) \\ E(e_t) \end{bmatrix}, \quad P_{a,0} = \begin{bmatrix} \text{var}(x_t) & 0 & 0 \\ 0 & \text{var}(v_t) & 0 \\ 0 & 0 & \text{var}(e_t) \end{bmatrix}, \quad (45)$$

and $\sqrt{P_{a,0j}}$ as the j th column of the Cholesky decomposition (say) of $P_{a,0}$. Given the diagonal form of $P_{a,0}$ (in this case), we have

$$\sqrt{P_{a,01}} = \begin{bmatrix} \sqrt{\text{var}(x_t)} \\ 0 \\ 0 \end{bmatrix}; \quad \sqrt{P_{a,02}} = \begin{bmatrix} 0 \\ \sqrt{\text{var}(v_t)} \\ 0 \end{bmatrix}; \quad \sqrt{P_{a,03}} = \begin{bmatrix} 0 \\ 0 \\ \sqrt{\text{var}(e_t)} \end{bmatrix}.$$

The seven columns of $X_{a,0}$ are then generated by

$$E(X_{a,0}); \quad E(X_{a,0}) + a_j \sqrt{P_{a,0j}}; \quad \text{for } j = 1, 2, 3; \quad E(X_{a,0}) - b_j \sqrt{P_{a,0j}}; \quad \text{for } j = 1, 2, 3,$$

where $a_1 = a_x$, $a_2 = a_v$ and $a_3 = a_e$, and the corresponding notation is used for b_j , $j = 1, 2, 3$.

2. Propagate the $t = 0$ sigma points through the transition equation as $X_{x,1} = k_1(X_{a,0}, \beta)$ and estimate the predictive mean and variance of x_1 as:

$$E(x_1|y_0) = \sum_{i=1}^7 w_i X_{x,1}^i \quad (46)$$

$$\text{var}(x_1|y_0) = \sum_{i=1}^7 w_i (X_{x,1}^i - E(x_1|y_0))^2, \quad (47)$$

where $X_{x,1}^i$ denotes the i th element of the (1×7) vector $X_{x,1}$ and w_i the associated weight, determined as an appropriate function of the a_j and b_j ; see Ponomareva and Date (2010).

3. Produce a new matrix of sigma points, $X_{a,1}$, for $t = 1$ generated by

$$E(X_{a,1}); E(X_{a,1}) + a_j \sqrt{P_{a,1j}}; \text{ for } j = 1, 2, 3; E(X_{a,1}) - b_j \sqrt{P_{a,1j}}; \text{ for } j = 1, 2, 3, \quad (48)$$

using the updated formulae for the mean and variance of x_t from (46) and (47) respectively, in the calculation of $E(X_{a,1})$ and $P_{a,1}$.

4. Propagate the $t = 1$ sigma points through the measurement equation as $X_{y,1} = h_1(X_{a,1}, \beta)$ and estimate the predictive mean and variance of y_1 as:

$$E(y_1|y_0) = \sum_{i=1}^7 w_i X_{y,1}^i \quad (49)$$

$$var(y_1|y_0) = \sum_{i=1}^7 w_i (X_{y,1}^i - E(y_1|y_0))^2, \quad (50)$$

where $X_{y,1}^i$ denotes the i th element of the (1×7) vector $X_{y,1}$ and w_i is as defined in Step 3.

5. Estimate the first component of the likelihood function, $p(y_1|y_0)$, as a Gaussian distribution with mean and variance as given in (49) and (50) respectively.
6. Given observation y_1 produce the up-dated filtered mean and variance of x_t via the usual KF up-dating equations:

$$E(x_1|y_1) = E(x_1|y_0) + M_1(y_1 - E(y_1|y_0))$$

$$var(x_1|y_1) = var(x_1|y_0) - M_1^2 var(y_1|y_0),$$

where:

$$M_1 = \frac{\sum_{i=1}^7 w_i (X_{x,1}^i - E(x_1|y_0))(X_{y,1}^i - E(y_1|y_0))}{var(y_1|y_0)}$$

and the $X_{x,1}^i$, $i = 1, 2, \dots, 7$ are as computed in Step 3.

7. Continue as for Steps 2 to 6, with the obvious up-dating of the time periods and the associated indexing of the random variables and sigma points, and with the likelihood function evaluated as the product of the components produced in each implementation of Step 5, and the log-likelihood produced accordingly.

C.2 Proof of Corollary 1

The result follows by verifying Assumptions **(A2)**-**(A5)** (Assumption **(A1)** can be verified by inspection) for the SV-SQ model.

From the auxiliary model in (22) and (23) we have that the transition density is given by

$$q_\beta(x, x') = \frac{1}{1 - \Phi(\varphi)} \frac{1}{\sqrt{2\pi\beta_3^2(x')^2}} \exp\left(-\frac{(x - \beta_1 - \beta_2 x')^2}{2\beta_3^2 x'}\right), \quad (51)$$

with $x \in X := \{x : x > \varphi + \beta_1\}$. This structure implies that

$$g_\beta(x, y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y - \ln(x))^2}{2}\right). \quad (52)$$

(A2) The stated restrictions on Φ guarantee the satisfaction of this condition.

(A3.1) Satisfaction follows from the definition of X , \mathcal{B} and the definition of $q_\beta(x, x')$ in (51).

(A3.2) To verify this condition, we have

$$\begin{aligned} \sup_{\beta \in \mathcal{B}} \sup_{x, x' \in X} q_\beta(x, x') &= \sup_{\beta \in \mathcal{B}} \frac{1}{1 - \Phi(\varphi)} \frac{1}{\sqrt{2\pi\beta_3^2}} \sup_{x, x' \in X} \frac{1}{x'} \exp\left(-\frac{(x - \beta_1 - \beta_2 x')^2}{2\beta_3^2 x'}\right) \\ &\leq \sup_{\beta \in \mathcal{B}} \frac{1}{1 - \Phi(\varphi)} \frac{1}{\sqrt{2\pi\beta_3^2}} \sup_{x' \in X} \frac{1}{x'} < \infty \end{aligned}$$

(A.4.1) Satisfaction follows from the definition of $g_\beta(x, y)$ in equation (52).

(A.4.2) Note that $\sup_{x \in X} g_\beta(x, y) = \exp(y) \exp(1/2)$. For any compact set \mathcal{K} ,

$$\lim_{x \rightarrow \infty} \sup_{y \in \mathcal{K}} \frac{g_\beta(y, x)}{\sup_{x' \in X} g_\beta(y, x')} = \lim_{x \rightarrow \infty} \sup_{y \in \mathcal{K}} \frac{\exp(-1/2)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y - \ln(x))^2 - y\right)$$

which is finite for all fixed $x > 0$, and any $\mathcal{K} \subset Y$ compact. The term converges to zero as $x \rightarrow \infty$ for any $y \in \mathcal{K} \subset Y$.

(A.4.3) From the definition $\ln^+(h(z)) = \max\{0, \ln(h(z))\}$ and $h(z) = \sup_{x \in X} g_\beta(x, z) = \exp(z) \exp(1/2)$, we have that

$$\mathbb{E}_\phi[\ln^+(h(z_0(\phi)))] = \mathbb{E}_\phi[z_0(\phi)] < \infty$$

for stationary distribution $f_0(z)$ with finite mean and $\sup_{z \in Y} f_0(z) < \infty$.

(A.4.4) This result follows similarly to **(A.4.3)** and is thus omitted.

(A5) Verification of **(A5)** depends on the specific structure of the filtering mechanism used to obtain the likelihood. To this end, we are required to explicitly incorporate the AUKF approach in construction of the likelihood. Define $X_{a,0}$ to be the (3×7) matrix of initial sigma-points, as referenced in Appendix C.1, where $X_{a,0}(j, i)$ is the element in the j -th row

and i -th column of $X_{a,0}$. For fixed weights $\{w_i\}_{i=1}^7$, with $\sum_{i=1}^7 w_i = 1$, $w_i > 0$, $i = 1, \dots, 7$, we initialize the system by propagating $X_{a,0}$ through the state equation. To build the auxiliary likelihood using the AUKF, then define the predicted mean and variance of x_t as

$$\begin{aligned}\hat{x}_{t|t-1} &= \sum_{i=1}^7 w_i k(X_{a,0}(1, i), X_{a,0}(2, i), \beta), \\ P_{t|t-1}^x &= \sum_{i=1}^7 w_i [k(X_{a,0}(1, i), X_{a,0}(2, i), \beta) - \hat{x}_{t|t-1}]^2, \\ k(X_{a,0}(1, i), X_{a,0}(2, i), \beta) &= \beta_1 + \beta_2 X_{a,0}(1, i) + \beta_3 \left(\sqrt{X_{a,0}(1, i)} \right) X_{a,0}(2, i),\end{aligned}$$

where the sigma points reflect the positivity of the variance. From $\hat{x}_{t|t-1}$ and $P_{t|t-1}^x$ the new matrix of sigma points, $X_{a,1}$, is produced. Define the predicted mean and variance for the observed y_t , based on the $X_{a,t-1}$ matrix of sigma-points, as

$$\begin{aligned}\hat{y}_{t|t-1} &= \sum_{i=1}^7 w_i h(X_{a,t-1}(1, i), X_{a,t-1}(3, i), \beta), \\ P_{t|t-1}^y &= \sum_{i=1}^7 w_i [h(X_{a,t-1}(1, i), X_{a,t-1}(3, i), \beta) - \hat{y}_{t|t-1}]^2, \\ h(X_{a,t-1}(1, i), X_{a,t-1}(3, i), \beta) &= \log(X_{a,t-1}(1, i)) + X_{a,t-1}(3, i).\end{aligned}$$

For $\zeta_t = y_t - \hat{y}_{t|t-1}$, the augmented Kalman filtering steps for x_t are as follows:

$$\begin{aligned}\hat{x}_{t|t} &= \hat{x}_{t|t-1} + M_{t|t} \zeta_t, \\ P_{t|t}^x &= P_{t|t-1}^x - M_{t|t}^2 P_{t|t-1}^y \\ M_{t|t} &= \frac{\sum_{i=1}^7 w_i [k(X_{a,t-1}(1, i), X_{a,t-1}(2, i), \beta) - \hat{x}_{t|t-1}] [h(X_{a,t-1}(1, i), X_{a,t-1}(3, i), \beta) - \hat{y}_{t|t-1}]}{P_{t|t-1}^y}.\end{aligned}$$

In accordance with the AUKF algorithm in the Appendix C.1, and noting the structure of the approximating model, the time- t matrix of sigma points $X_{a,t}$ is given by

$$X_{a,t} = \begin{pmatrix} \hat{x}_{t|t} & \hat{x}_{t|t} + a_1 \sqrt{P_{t|t}^x} & \hat{x}_{t|t} & \hat{x}_{t|t} & \hat{x}_{t|t} - b_1 \sqrt{P_{t|t}^x} & \hat{x}_{t|t} & \hat{x}_{t|t} \\ \lambda^* & \lambda^* & \lambda^* + a_2 \sqrt{\text{var}(v_t)} & \lambda^* & \lambda^* & \lambda^* - b_2 \sqrt{\text{var}(v_t)} & \lambda^* \\ \gamma^* & \gamma^* & \gamma^* & \gamma^* + a_3 \sqrt{\pi^2/2} & \gamma^* & \gamma^* - b_3 \sqrt{\pi^2/2} & \gamma^* \end{pmatrix}, \quad (53)$$

where $a_1 = a_2 = a_3 = b_1 = b_2 = b_3 = \sqrt{3}$, $\gamma^* = -1.27$,

$$\lambda^* = \frac{\phi\left(\frac{-\beta_1}{\beta_3}\right)}{1 - \Phi\left(\frac{-\beta_1}{\beta_3}\right)} \text{ and } \text{var}(v_t) = \left[1 - \lambda\left(\frac{-\beta_1}{\beta_3}\right) \left(\lambda\left(\frac{-\beta_1}{\beta_3}\right) - \frac{-\beta_1}{\beta_3} \right) \right]. \quad (54)$$

Using the definitions $\hat{y}_{t|t-1}$, $P_{t|t-1}^y$, ζ_t , the auxiliary log-likelihood, conditioning on y_1 , is given by

$$L_a(\mathbf{y}; \beta) = \sum_{t=2}^T \ell(y_t | \mathbf{y}_1^{t-1}; \beta) = - \sum_{t=2}^T \ln(P_{t|t-1}^y) + \frac{1}{2} \frac{\zeta_t^2}{P_{t|t-1}^y},$$

where a Gaussian approximation for the components of the likelihood function is adopted at this point.

Primitive conditions guaranteeing **(A5)** are as follows: 1) $\mathbb{E}_\phi[|\ell(z_t(\phi); \mathbf{z}_1^{t-1}(\phi), \beta)|] < \infty$ for all $\beta \in \mathbf{B}$, where $L_a(\mathbf{z}(\phi); \beta)/T = \frac{1}{T} \sum_{t=1}^T \ell(z_t(\phi); \mathbf{z}_1^{t-1}(\phi), \beta)$; 2) $\ell(z_t(\phi); \mathbf{z}_1^{t-1}(\phi), \beta) \neq \ell(z_t(\phi); \mathbf{z}_1^{t-1}(\phi), \tilde{\beta})$ for all $\beta \neq \tilde{\beta}$. Condition 1) is satisfied by an extension of Proposition 10 part (i) in Douc and Moulines (2012) to the case of stationary $z_0(\phi)$; these details are omitted for brevity but are available from the authors upon request. For Condition 2) to be satisfied, the AUKF recursions must be unique in β . Uniqueness of the AUKF recursions requires that the matrix of sigma-points be unique in β for each $t \geq 1$. Denote by $X_{a,t}(\beta)$ the (3×7) matrix of sigma-points in (53) constructed for a given β . Focusing on the elements of $X_{a,t}(\beta)$ due to $\hat{x}_{t|t}$, by the Kalman recursions for this model, $\hat{x}_{t|t}$ is a unique function of β and so $X_{a,t}(\beta) \neq X_{a,t}(\tilde{\beta})$ if $\beta \neq \tilde{\beta}$, and the result follows.⁸

⁸We focus on the elements within the Kalman recursion portion of $X_{a,t}(\beta)$, since $\lambda^*(\beta)$ is not one-to-one in the parameters β_1, β_3 and so there exists $\tilde{\beta} \neq \beta$ such that $\lambda^*(\beta) = \lambda^*(\tilde{\beta})$.